nature

ANTARCTIC BIODIVERSITY

The Southern Ocean fauna yields its secrets

nature

www.nature.com/nature

Vol 447 | Issue no. 7142 | 17 May 2007

Never had it so good?

The Blair–Brown era has been a golden one for British science.

Anyone remember Save British Science? Twenty-one years ago, with universities in decline, laboratories in disrepair and academics packing their bags to head in the general direction of California, there was plenty of meat for the pressure group of that name to sink its teeth into. Today, the shrill sound of the name seems almost quaint.

No one would claim that British scientists today live in the best of all possible worlds, but there is no denying that, in many disciplines and departments, British science is in rather good shape, thank you very much. The country's universities have their problems but do well in international comparisons; research-intensive industries, such as pharmaceuticals, are prospering; and scientific leaders have a respected voice in national affairs. In 2005, Save British Science accordingly transformed itself into the rather more staid-sounding Campaign for Science and Engineering.

This happy state of affairs is one success for which Tony Blair, Britain's outgoing prime minister, and Gordon Brown, set to be his successor, can share credit — a success that the political world has not fully appreciated. Blair is not viscerally grabbed by science in the same way as Al Gore, but he is attracted by its inherent modernity, and has been steadfast in supporting Brown, the chancellor of the exchequer, in his deeper devotion to the issue.

Brown sees science and innovation at the very core of his national modernization project. As the Labour Party's spokesman on trade and industry soon after Save British Science was set up, he read, absorbed and planned. The outcome was the decision by 'New Labour', which he and Blair brought into being, to ditch the party's traditional attachment to applied research and industrial policy, and concentrate instead on backing basic science while creating a better climate for entrepreneurial innovation.

Blair announced on 10 May that he was resigning; on 27 June, barring acts of God, Brown will succeed him. (The bookmaker William Hill, to which *Nature* regularly turns on matters of political prognosis, is already paying out to punters with money wagered on Brown's ascendancy.) For science, this should mean more of the same. Brown has made it crystal clear that the government's commitment to investment must continue. One of his last acts at the Treasury has been to publish a ten-year spending plan that foresees substantial further growth in research spending.

However, there are two notable concerns on the horizon. One is that, faced with day-to-day responsibility for everything from hospital waiting lists to the occupation of Iraq, Brown is bound to leave decisions about science and innovation in the hands of colleagues. Few senior Labour politicians other than Brown have shown much interest in, or commitment to, science, so this change could bode ill.

Second, Brown's electoral prospects are less bright than those of Blair before the past three elections. Brown leads a party with which the electorate is disenchanted against a Conservative opposition that has found new appeal under the leadership of David Cameron. There is scant indication that the Conservatives are champing at the bit to invest money in the universities. Instead, they are likely to seek reductions in public spending in almost all areas that are less politically sensitive than health or schools — areas such as scientific research.

This puts a new onus on British researchers. Their current happy state is not of their own making. With the plucky and laudable exception of Save British Science, the research community has done little to improve its lot. Now it should give its support to the renamed Campaign for Science and Engineering, and do what it can to maintain the happy status quo.

"There is scant indication that the Conservatives are champing at the bit to invest money in the universities."

Researchers should make it a priority to communicate their accomplishments to politicians of all parties and to the public at large, thus justifying the generosity they have received. The scientists closest to the levers of power should identify those among their number best suited to the key jobs, such as director-general of the research councils, that can exert real budgetary influence. And the community as a whole should be building alliances — with industry, consumers (especially of healthcare), environmentalists and everyday geeks — that will sustain political support for science in the long term. British science has had the good fortune to be saved by far-sighted politicians. But its future lies in its own hands. ■

Health cheques

Philanthropy offers a valuable approach to funding.

In scientific funding, as in agriculture, monoculture is risky. However well-meaning and diligent its paymasters may be, a field that has only one source of funding runs the risk of missing opportunities and succumbing to a shared perception of what is possible and what is not. Diversity is thus to be prized. The current increase in scientific funding from individual philanthropists, private foundations and non-profit organizations, particularly in biomedical science (see page 248 and www.nature.com/news/specials/philanthropy), is a welcome development in a field that is largely dominated by governments.

With this new money come new attitudes. Philanthropists tend to have strong and clear ideas about what should be funded. If, as many of the new wave did, they made their money in industries that are themselves driven by research, they will often want to delve deeper into what scientists are actually doing than the foundations

of yesteryear, either out of pure intellectual engagement or a desire to micromanage.

Philanthropic organizations tend to be interested in particular outcomes, rather than in simply funding the most academically interesting question that presents itself. They may take a more business-like approach, too, demanding milestones and results. And they can show a greater readiness to turn off the funding if these things are not forthcoming — or, in some cases, simply if the stock market takes a tumble.

The mixture of risk tolerance and accountability seen in some foundations, particularly those concerned with a specific disease, may seem contradictory, but it need not be so. What matters to the administrators who dole out the cash is normally not success per se, but demonstrated diligence in following the risky strategy set down. This level of accountability can make some researchers uncomfortable (see page 252). If basic researchers expect to benefit from these disease-focused foundations, they need to make a compelling case that their studies offer a real chance of progress towards a cure, and to document the way those chances play out.

This does not mean that the donor is always right, however. There are real risks that some areas of biomedical research could be dominated by a few rich people and the boards on which they sit, with little guidance from scientific consensus or expertise. Philanthropic groups need to think carefully about how best to award money, and to realize that no single strategy will succeed in every case. What's right for Huntington's disease, where a faulty gene has been named and shamed, may be wrong or premature for something more complex, such as autism. Rushing into the wrong strategy will merely waste time and money.

In the end, though, it is the philanthropists' money to waste. They would obviously be wise to make sure that the money is well spent, but they cannot always be stopped from other courses. Rather, it is for the scientists who stand to benefit from such largesse to ask themselves whether the money on offer is worth the risks of the course being prescribed. By doing so, and by helping donors identify the most pressing scientific questions relevant to their remit, the scientific community can help make the most of the new diversity offered by the philanthropy boom. ■

Blurred vision

In the end, the European Institute of Technology will not be worthy of its title.

It could have been inspiring. Imagine a university set up to educate some of the best students from Europe and beyond, and to provide a home for the world's most creative researchers in the natural sciences and engineering. It would be an institution with an outstanding new campus, benefiting from annual revenues of at least €500 million (US$700 million) — a pinnacle scholars would aspire to and entrepreneurs would move to be close to.

That is what the European Institute of Technology (EIT) might have been. It was what was implicit in the idea that the president of the European Commission, José Manuel Barroso, put forward under that name in 2005. The proposal that now goes by the name of the EIT is a paltry thing by comparison — a small, central, administrative executive that would select 'knowledge and innovation communities' (KICs) to fund research in promising-sounding areas. The KICs would be distributed networks made up of 'partnerships' between organizations in the education, research and business sectors, and would confer postgraduate degrees. The commission would contribute €2.4 billion between 2008 and 2013.

An expert group commissioned by the European Parliament to analyse this proposal recommended a different way forward last month. In this vision, too, there is no central EIT — but the KICs become genuine bricks-and-mortar institutes, 20 or so of them, each with 300-odd scientists, located in regions where there is already established research strength.

Parliament, however, seems not to be taking its experts' advice. Both it and the European Council — the two decision-making bodies of the European Union — are likely to approve the commission's lack-lustre notion of virtual networks. And both are downplaying the element of education, toying with the idea that the EIT and its components might offer some sort of watered-down postgraduate diploma without the clout of a PhD.

So Europe will get yet another virtual industry–academic network, to sit alongside the Framework Programme's Networks of Excellence, the EUREKA clusters and a welter of other variations on the theme. This shop-worn notion of the EIT has little if any support among researchers. Industry, too, finds it uninteresting.

This delocalized compromise marks a distressing loss of nerve. Europe has a demonstrated ability to deliver enduring world-class research institutions: the particle-physics laboratory CERN and the European Molecular Biology Laboratory stand as evidence. It has shown a willingness to invest in research excellence for its own sake, independent of political agendas — witness the fledgling European Research Council, flooded with applications in its first round of awards. If it chose to, why should the world's largest economy not set up a unitary engineering powerhouse to rival the Massachusetts Institute of Technology, with all the benefits that such scale and ambition confer?

This pitiable state of affairs highlights much of what is wrong with pan-European politics. One fundamental problem is that of 'subsidiarity', a principle embedded in the European Union's foundational Treaty of Rome that requires the commission to steer clear of functions that individual countries are considered to do well enough already, including the award of academic qualifications. Another is a chronic reluctance to commit to bricks and mortar for research.

The German presidency is keen to launch a 'test KIC' next year, specializing in energy efficiency or climate change. It will be a long way from a world-leading academic institution that would have inspired Europe's young people and stimulated its economies. The financial and constitutional challenges involved would have been undeniably great. But the outcome could have been fully worthy of the academic heritage that Europe possesses, and may have stimulated it to new heights. Is that vision now dormant? Or is it dead? ■

RESEARCH HIGHLIGHTS

STEM CELLS

Becoming a blood vessel

Nature Meth. doi:10.1038/nmeth1041 (2007)

Stem-cell therapies have enormous potential to replace defective blood vessels, such as those destroyed by heart attacks and degenerative eye disease.

Now, Robert Lanza of the company Advanced Cell Technology in Worcester, Massachusetts, and his colleagues have made progress towards coaxing stem cells into becoming the right kind of cells for the job. The group reports a two-step method for making large amounts of regenerative cells, called haemangioblasts, from human embryonic stem-cell cultures. The team shows that when injected into rodents, these haemangioblasts give rise to cells that improve symptoms of heart attack, eye disease and stroke. The method is more efficient than those previously reported, and its well-defined protocol should be reproducible by other labs.

ECOLOGY

Urban chicks

Ecology **88,** 882–890 (2007)

City birds migrate less frequently than country birds, and their breeding patterns have changed.

Jesko Partecke and Eberhard Gwinner (deceased) of the Max Planck Institute for Ornithology in Andechs, Germany, have found a link between these two life-history traits. They kept hand-raised European blackbirds (*Turdus merula,* pictured below), a partially migratory species, from both urban and forest populations under identical conditions for two years.

Compared with their country cousins, they found that urban males, but not females, showed less of two principle features of a migratory disposition: night activity and fat deposition. The researchers suggest that there may be an evolutionary benefit for the urban males to winter at home, so that they can access prime breeding habitats and start mating before migratory birds return.

M. DANEGGER/NHPA

Heart-disease hunt

S. CAMAZINE & S. TRAINOR/SPL

Science doi:10.1126/science.1142447 and doi:10.1126/science.1142842 (2007)

Disease-gene hunting, anyone? It's open season. Following a slew of reports linking common single-letter DNA variations to diseases ranging from diabetes to Crohn's, researchers are closing in on the world's biggest killer.

Two independent groups have identified single-nucleotide polymorphisms, or SNPs, that confer a significantly increased risk of heart disease in white people. All reside in the same tight neighbourhood on chromosome 9. They aren't found in a gene, but in a non-coding region that abuts two tumour-suppressor genes and that scientists speculate may regulate gene expression or function.

Ruth McPherson at the University of Ottawa Heart Institute in Canada and her colleagues used SNP-seeking gene chips to identify one cluster of SNPs. In studies of the DNA of 23,000 Caucasians, those subjects who carried this high-risk cluster on both of their chromosome 9s were 30–40% more likely to have heart disease than those without it. The X-ray pictured shows the chest and heart of a healthy person.

Kari Stefansson of deCode Genetics in Reykjavik, Iceland, and colleagues found an even stronger association in some 17,000 Caucasians. Those who carried two copies of the SNP that this team identified in the same region of chromosome 9 had a 64% higher risk of heart attack than those subjects who did not carry the variation — and more than twice the risk of having a heart attack young.

NANOMEDICINE

This won't hurt a bit

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0700567104 (2007)

The world's thinnest hypodermic needles have been made from carbon nanotubes just 10–20 nanometres wide. Carolyn Bertozzi, Alex Zettl and their co-workers at the University of California, Berkeley, have used these nanoneedles to inject material into live human cells, leaving no visible damage to the cell membranes.

The team attached tether molecules to the nanotube walls that contain short molecular chains tipped with biotin, an organic molecule that binds tightly to the protein streptavidin. These served as hooks for streptavidin-coated fluorescent nanoparticles, which coated the nanotube surface. When the needle, attached to the tip of an atomic force microscope, is pushed through the cell membrane, the tether is chemically snipped and the nanoparticle cargo released.

BIOCHEMISTRY

A bacterial nose

J. Am. Chem. Soc. doi:10.1021/ja0692480 (2007)

A bacterial cell capable of following a particular chemical trail could track down pollutants or help to combat disease — especially if equipped with machinery for degrading the source of the 'smell'.

Shana Topp and Justin Gallivan at Emory University in Atlanta, Georgia, have designed such a general-purpose search function for *Escherichia coli*. Because re-engineering the receptor proteins involved in *E. coli*'s natural chemical-based movement (chemotaxis) is difficult, the researchers pursued another strategy — they designed a way to immobilize any bacteria that go off target.

The team made RNA switches sensitive to a particular signal or 'smell' — here a chemical very similar to caffeine. If there is none of this chemical about, the switch will prevent production of a protein needed for 'swimming'. But the presence of the signal flips the switch and activates the production of the protein, enabling the cells to, in effect, swim up a gradient of increasing signal concentration — although many immobilized *E. coli* that went in the wrong direction will litter the route to the source.

GEOLOGY

Seasonal shaking

Geophys. Res. Lett. **34,** L08304 (2007)

There are more earthquakes in the Himalaya in the winter than in the summer, according to a new study of Nepalese seismic records from 1995 to 2000.

The fact that there are 37% more earthquakes in winter in Nepal, say Laurent Bollinger of the French Atomic Energy Commission and colleagues, can be attributed, indirectly, to the weather. The researchers theorize that summer monsoons weigh down northern India with rainfall and fill underground aquifers. This changes the stress patterns in the surrounding rock and in the fault systems that stretch up into the Himalaya, suppressing summer seismicity.

ECOLOGY

Stream of tears

Glob. Clim. Change Biol. **13,** 942–957 (2007)

Although many studies show the grim ecological effects of climate change, some ecosystems are somewhat neglected, including streams. Isabelle Durance and Steve Ormerod of Cardiff University, UK, have shown, through a 25-year project on Welsh stream macroinvertebrates such as

CARDIFF UNIV.

mayflies and stoneflies, just how big the effects might be. After the team factored out other cyclical climate phenomena, they concluded that winter temperatures in the streams rose by 1.4–1.7 °C during the study period, with clear ecological consequences.

Their results suggest that in the most species-rich streams the abundance of invertebrates in the spring-time could decline by one-fifth for every degree of temperature gain. An increase of 3 °C could see up to ten species extinct locally — as much as 25% of the typical richness, the scientists say.

MOLECULAR BIOLOGY

All about growing yeast

J. Biol. **6,** 4 (2007)

Researchers have catalogued the changes in gene expression, protein abundance and metabolite composition that accompany cell growth in the yeast *Saccharomyces cerevisiae*.

Previous attempts to characterize cell growth did not separate the effects of growth from responses to nutrient depletion as the growing cells exhausted nutrients from the medium. Now, Stephen Oliver of the University of Manchester, UK, and his colleagues have grown yeast cultures under four different nutrient limitations and at three different growth rates. They then measured changes in the abundance of thousands of cellular compounds, looking for trends dependent on growth rate and independent of nutrient deficiency.

Their results provide a catalogue of parameters that could be used in mathematical models and to make genetic engineering of metabolic pathways easier.

CHEMISTRY

Snap, crackle and glow

J. Am. Chem. Soc. doi:10.1021/ja0716498 (2007)

Bubbles created by ultrasound can collapse violently and agitate crystals of certain chemicals to the point of bursting. These crystals then give off light. And the gases released in the process can react with each other, say researchers at the University of Illinois in Urbana-Champaign.

Kenneth Suslick and Nathan Eddingsaas closely watched this mechanoluminescence in a slurry of the organic molecule resorcinol in the solvent dodecane. The emitted energy responsible for the luminescence was 1,000 times as intense as that produced if the same crystals had been cracked by mere milling or grinding. Suslick attributes this to the increased speed of molecular collisions under ultrasound conditions, and the subsequent increase in emitted energy from both inert and radioactive gases caught in the discharge. It is those gases, specifically oxygen and small hydrocarbons, that went on to react.

JOURNAL CLUB

Eric Wolff
British Antarctic Survey, Cambridge, UK

An ice-core scientist wonders what makes the Earth run hot and cold.

In the past 800,000 years, Earth has seen long, cold phases punctuated every 100,000 years by short, warm interglacials. If I claim to understand climate, then I should know why these cycles occur and why we are in a warm phase today.

The most obvious external controls on our climate are small changes in Earth's orbit. These affect the variation of incoming sunlight (insolation) with season and latitude. 'Milankovitch theory' says that this in turn controls the occurrence of glaciations.

There is one obvious problem: although 100,000 years is the period of eccentricity of Earth's orbit, insolation shows much stronger effects at shorter periods, such as 41,000 and 23,000 years.

A recent paper (E. Tziperman *et al. Paleoceanography* **21,** PA4206 doi:10.1029/2005PA001241; 2006) suggests a way around this. It uses a model in which climate varies with an average period controlled by internal features — such as the time needed for ice-sheet growth — on a 100,000-year timescale.

However, the exact timing of climate changes is paced by orbital cycles at shorter periods. The result is that a wide range of plausible internal controls on climate can give similar predictions of how climate has evolved with time, all of them with a 'Milankovitch imprint'.

This frees us from the apparent misconception that we need an external forcing with a period of 100,000 years, but it does not identify the internal mechanisms responsible.

I used to think this was a problem for others to solve, but as part of the team that extended the ice-core record back 800,000 years, I have the tantalizing hope that the clues we need might be locked in our cold room.

NEWS

United Nations shelves action on indoor hazards

A United Nations (UN) meeting has failed to agree on an action plan to deal with indoor air pollution — the range of hazards related to cooking indoors that is thought to kill more people every year in poor countries than malaria.

Despite two weeks of negotiations in New York, the UN Commission on Sustainable Development was unable to ratify a draft communiqué on indoor pollution and other developmental issues that were up for discussion. But activists say the very fact that the discussion took place represents valuable progress in acknowledging the scale of the pollution problem.

"The problem of indoor air pollution has been basically invisible."

At the end of a fractious meeting that culminated in the election of the Zimbabwean environment minister, Francis Nhema, to chair the commission, representatives from Switzerland and the European Union (EU) rejected the draft communiqué, saying that its vacuous content would threaten past agreements and contained no goals that would spur action on a number of key issues.

"In previous discussions, the problem of indoor air pollution has been basically invisible, perhaps because it is a situation that affects mainly women," says Maria Arce Moreira, a policy adviser to Practical Action, a British pressure group that works on poverty issues. "It's important that it is finally recognized as a problem, but the proposed actions to deal with it are not enough."

Around half of the world's population cooks on stoves that burn biomass such as wood, crop residues or dung, development specialists say. According to estimates from the World Health Organization (WHO), smoke emitted by traditional cookers kills 1.6 million people each year, most of them women and children (WHO *Indoor Air Pollution: National Burden of Disease Estimates*; 2007).

Researchers predict that if these trends continue, in Africa alone indoor air pollution will kill 10 million people by 2030. Of these deaths, up to 3.7 million could be saved by switching to petroleum-based fossil fuels such as kerosene (R. Bailis *et al. Science* **308,** 98–103; 2005).

Lung cancer, pneumonia and acute lower-respiratory infections are prevalent as a result of constant exposure to carbon monoxide, particulates, hydrocarbons and carcinogens such as formaldehyde and benzene that are contained in cooking smoke.

Activist groups, including Practical Action, have previously called on governments to adopt firm measures to halve the number of people cooking with traditional fuels. But with fossil-fuel prices at historic highs, and most of the world's poorest people using wood-burning stoves, there is little appetite for such measures.

Nonetheless, EU representatives wanted the UN commission to ask nations and regions to set appropriate targets, "because without targets you cannot easily review these issues", says Natascha Beinker, a policy adviser at the

Modellers seek reason for low retraction rates

Computer scientists at Columbia University in New York have used a mathematical model to estimate the number of flawed scientific papers that go unretracted, and its relation to journal impact factors.

In correspondence published in *EMBO Reports* (M. Cokol *et al. EMBO Rep.* 5, 422–423; 2007), the researchers find that fewer papers are retracted by journals with low impact factors. But their model raises as many questions as it answers, say specialists in scientific publishing, some of whom argue that it greatly oversimplifies the issues.

Murat Cokol and his colleagues at the biomedical-informatics department at Columbia downloaded data for 9.4 million articles published between 1950 and 2004 from PubMed, an index of biomedical and general scientific literature. They identified 596 retracted articles — flagged up as such in PubMed — and found some striking relationships between the numbers of retractions and the impact factors of the journals that had published them.

Journals with high impact factors retract more papers, and low-impact journals are more likely not to retract them, the study finds. It also suggests that high- and low-impact journals differ little in detecting flawed articles before they are published.

Finally, the authors ran a model to estimate how many articles should have been retracted, and came up with 10,000 in a best-case scenario and more than 100,000 in a worst-case one. Most of the papers that needed to be retracted were published in low-impact journals.

The Cokol study was not peer-reviewed. Aviv Bergman, director of the Center for Computational Genetics and Biological Modeling at Stanford University in California, says that the researchers' modelling techniques are sound, but that he isn't in a position to judge their input data.

But scientists and editors familiar with retraction issues are sceptical of the quality of the model's input data. Theoretical modelling exercises will generate bad results if the input data are flawed, says Drummond Rennie, deputy editor of the *Journal of the American Medical Association*, and a medical researcher at the

J. BOETHLING/STILL PICTURES

Cooking with fire: wood-burning stoves are a major source of indoor pollution.

German ministry for economic cooperation and development, who attended the meetings.

EU representatives also called for a commitment to switching to cleaner biofuels, and to delivering financial aid and addressing the health risks from indoor cooking, "but there was not total commitment for this", Beinker adds.

"The Americans and Australians seem less willing to accept that there is a policy issue," says Andrew Scott, policy director for Practical Action. African and South American nations, which see environmental rules as restraints on their economic development, agreed that the issue was not a political one and blocked the EU move.

Peter Davies, an energy adviser to the UK Department for International Development, who was also at the meetings, argues that progress remains possible without firm targets. "Goals and targets are not something developed countries can push on. These are issues that national governments need to decide for themselves with their own national-development and poverty-reduction plans."

This year's meeting of the sustainable-development commission concentrated on four main issues — energy, industrial development, air pollution and climate change. The most contentious debate was fuelled by discussions on acceptable sources of energy.

"The EU and First World groups tended to be influenced by environmental issues, whereas the G77 [the developing nations] seemed to take a much harder-nosed look at what they need for their economic development," says Davies.

The lack of consensus means that rather than producing a final summary agreed by all countries, this year's chair, Abdullah bin Hamad al-Attiyah of Qatar, will issue a text that merely lists the points discussed.

"The final text will be very watered down now," says Arce Moreira. "We do not envisage support for energy issues related to indoor air pollution, such as addressing access to modern energy for the poor."

But Davies says that what matters is that the hitherto obscure issue of indoor pollution is gradually raising its political profile. "Whatever the chair's summary says will be pretty bland. It has been negotiated to the lowest common denominator. But I think it will be a mistake to look at the chair's summary and say this is it," he says. "The process and the debate it generates has value in itself." ■

Narelle Towie

SCORE

BURNING WOOD TO POWER FRIDGES
A low-pollution domestic device for the developing world.
www.nature.com/news

Institute for Health Policy Studies, at the University of California, San Francisco.

Although the number of retracted articles is probably only the tip of the iceberg in terms of the number that should have been retracted, the model — based on journal impact factor and number of retractions — is too simplistic to capture the complex reality of the issues affecting the size and nature of the hidden part, Rennie says.

The model clumps data from 1950 to 2004, for example, whereas trends are likely to be affected by the fact that the United States first introduced official policies on research misconduct twenty years ago, and that other nations did so even later, says Rennie.

Experience also shows that retraction figures are skewed by the fact that once misconduct is detected in one article by a researcher, dozens of articles by the same author often need to be retracted. "Digging into the data behind all these other articles is a truly monumental task, but until it's done, no one has much clue what the real number of retractions should be," says Rennie.

"No one has much clue what the real number of retractions should be."

Impact factor alone is also a very broad yardstick, says one scientific-literature specialist at PubMed, who points out that the impact factors of individual articles vary widely, even when they are in the same journal. Other models could also be made to fit the data with potentially very different outcomes, he says. High-impact journals might attract flawed papers, he speculates, simply because they publish cutting-edge research, in which competition and time pressure may favour both errors and misconduct.

In the correspondence, Cokol argues that the larger number of retractions in high-impact journals reflects the fact that they receive more scrutiny. But Sandra Titus, director of intramural research at the US Office of Research Integrity in Rockville, Maryland, says that's too simple a verdict. "It's only part of the issue," she says, adding that "legal barriers to retraction are so awkward that many journals simply pass rather than face the hassle."

Cokol defends his approach as a valid one to start exploring data on retraction. "All models are wrong, but some are useful," he says. "Our model certainly does not capture the reality in full, and no model does. But it captures certain aspects and gives a general direction on how to understand the issue better." ■

Declan Butler and Jenny Hogan

Purdue dogged by misconduct claims

Purdue University has set up a third panel to look at allegations of research misconduct against nuclear engineer Rusi Taleyarkhan. The existence of the latest inquiry emerged when a congressional committee wrote to the university questioning the thoroughness of its earlier review. Now Purdue, based in West Lafayette, Indiana, faces an uphill task in convincing the wider scientific community that it is mounting a serious investigative effort, critics say.

Since 1999, Taleyarkhan has received more than $780,000 from the Defense Advanced Research Projects Agency (DARPA) for work towards the goal of 'bubble fusion', in which energy-releasing fusion reactions are triggered by passing ultrasound through deuterated liquids[1]. In February, Purdue cleared him of misconduct after two inquiries[2,3]. But a third panel was set up in March following complaints to the inspector-general of the Office of Naval Research, which oversees DARPA grants.

Its existence came to light on 10 May, in a memo from the investigations and oversight subcommittee of the House Committee on Science and Technology, chaired by Brad Miller (Democrat, North Carolina). "I sincerely hope that the next inquiry will be conducted in a manner worthy of your great institution," Miller wrote to Purdue president Martin Jischke.

A subcommittee-staff memo sent with his letter alleges that Purdue's previous efforts on the issue were "not thorough", "never addressed the validity of the underlying research" and "did not follow" established procedures for investigating misconduct allegations. The criticisms are partly based on the subcommittee's analysis of internal Purdue documents that it requested in March.

According to the memo, these documents show that in March 2006, Purdue charged a panel with determining the facts surrounding articles published in *Nature* and elsewhere that reported concerns over the validity of Taleyarkhan's research. This first panel reviewed a limited number of documents, interviewed only two professors inside Purdue, and concluded its work by recommending further interviews and examination of lab notebooks, the memo says. University officials then set up a second inquiry panel that, according to the memo, did not seem to follow the recommendations of the first, but instead solicited and examined a narrower set of allegations.

PURDUE UNIV.

A third panel will probe bubble-fusion findings at Purdue University.

The memo reports that the second panel concluded that Taleyarkhan showed "what might be characterized most favorably as severe lack of judgment" when he participated in the preparation of a manuscript by a postdoc and a graduate student that claimed positive results for bubble fusion and did not name him as an author. Taleyarkhan then cited the work as "independent" confirmation of his own earlier research claims. The claim would not be accepted by the wider scientific community, the inquiry found. The memo says that the inquiry found Taleyarkhan had "abused his privilege as senior scientist" and placed the junior scientists in "precarious positions". But the inquiry concluded that no full-scale investigation into the possibility of misconduct or the underlying research was warranted and the university then cleared Taleyarkhan of wrongdoing in a press release.

Immediately after the release of Miller's letter and the accompanying memo, Purdue issued a statement saying that although it "could debate" some of its contents, the university would instead focus on its fresh inquiry.

Taleyarkhan's long time co-author, Richard Lahey of Rensselaer Polytechnic Institute in Troy, New York, took issue with the memo, saying that he had full confidence in Taleyarkhan's abilities and integrity and that the true story was one of "an exciting scientific discovery". He criticized congressional staff members for failing to contact him or Taleyarkhan's other co-authors as a part of their own inquiries.

Purdue's new inquiry is understood to be covering some of the ground omitted by earlier ones, including allegations that Taleyarkhan's work may be fraudulent, a concern first received by the university in June 2006. The new panel consists of unnamed panellists from the previous inquiries, because, Purdue says, these people are "familiar with the issues". In its statement, Purdue pledged to add one or more outside scientists to the panel.

India struggles to find director for top research agency

India's largest research agency has been without a permanent director for almost five months, prompting some observers to claim that science isn't high on the government's agenda.

The role of director of the Council of Scientific and Industrial Research (CSIR), which runs 37 laboratories staffed by more than 18,000 scientists, comes with the status of a government secretary. It became vacant last December when chemist Raghunath Mashelkar quit the post — and the hunt for a successor has got nowhere fast.

Science minister Kapil Sibal initially appointed Visveswaraiah Prakash, director of the CSIR's Central Food Technological Research Institute in Mysore, but he withdrew on health grounds. Maharaj Kishan Bhan, the government's biotechnology secretary, took temporary charge in January, but after eight weeks he passed the responsibility to Thirumalachari Ramasami, secretary of the Department of Science and Technology.

India's prime minister, Manmohan Singh, keeps saying the government gives priority to science, says Sri Krishna Joshi, director of the CSIR from 1991 to 1995. "But one does not get the feeling he is serious, considering that the CSIR, of which he himself is president, has been headless for so long."

Scientists blame the declining prestige of public research as one reason for the failure to appoint a director. "Until a few years ago, the only way for a scientist to get noticed was by becoming a secretary to a scientific department," says Samir Brahmachari, who heads

But that may not be enough to restore credibility in the university's process, says C. K. Gunsalus, a lawyer at the University of Illinois at Urbana-Champaign who specializes in research misconduct. "They need to completely reconstitute the process. Different panellists, including at least one external to the institution, are required to restore faith in their findings. It's the only way to close the matter."

Gunsalus says the case raises a broader issue of how officials respond to misconduct accusations. "There's an unwillingness to be embarrassed," she says. "People can make intuitive, but mistaken, decisions counter to the underlying interest, which is the integrity of the institution."

Purdue says that it will continue to adhere to its own policy and to federal guidelines. Responding to Purdue's statement, Taleyarkhan said: "I have confirmed my full and complete cooperation with the new inquiry to be introduced this year at Purdue and, in fact, welcome the opportunity to once again clear up the doubts raised in the press."

At Purdue, nuclear-engineering faculty expressed mixed feelings about the affair. Chan Choi, who chaired a fact-finding internal committee that looked into the authorship issues, says that institutional blunders should not overshadow the ample sound research performed at the university. "I think they have to make the charges clear and I think the committee's names should be known," he says. But he adds that the outside interest has provided an opportunity for the university to improve its procedure now: "That's a healthy message from Congress." ■

Eugenie Samuel Reich

1. Taleyarkhan, R. P. *et al. Science* **295,** 1868–1873 (2002).
2. Reich, E. S. *Nature* **445,** 690–691 (2007).
3. Reich, E. S. *Nature* **446,** 480 (2007).

the CSIR's genomics institute in New Delhi. "Today, with opportunities to do good science outside the government, no brilliant scientist would want to be a secretary and be answerable to the parliament and all sorts of committees."

Scientists at the CSIR agree that it's a difficult job. During Mashelkar's eleven years as director, unions were banned and complaints from staff were rarely entertained. "The new director should be ready to deal with the thousands of complaints that have piled up," says one lab director who wants to remain anonymous. ■

K. S. Jayaraman

D. FLANNIGAN & K.S. SUSLICK, UNIV. ILLINOIS, URBANA-CHAMPAIGN

BUBBLE FUSION
Find the back-story of Taleyarkhan's table-top fusion research online.
www.nature.com/news/bubblefusion

Anger at 'unfit' museum design

The directors of the Natural History Museum in London are under fire over plans to split parts of its world-renowned collection of biological specimens on a permanent basis. Some of the museum's curators are angry that plans for a new building to store the bulk of the museum's entomology and botany collections remain unaltered three years after being branded unfit for the purpose by museum staff.

Work has now started on the £73-million (US$140-million) building for the Darwin Centre Phase Two, even though there won't be space to house all of the museum's 28 million insect specimens and 5.5 million plants. The structure will replace the previous Entomology Building, which was demolished in 2006: it had been impregnated with potentially cancer-causing naphthalene, which helped to preserve specimens.

"Our first gripe is that they wilfully destroyed a sound building without looking at the alternatives," says Henry Barlow, an entomologist who contributed some 30,000 specimens to the museum's collections and a member of Friends of the Natural History Museum, a group of donors and visitors that liaises with museum staff. Barlow accuses the museum of being more interested in winning architectural awards than properly curating its collections.

Museum directors found that refurbishing the previous building was not cost-effective, says the museum's head of collections, Mike Fitton. He says when the new building opens in 2009, it will offer better conditions for specimens, better research facilities and improved public access, so visitors can see the museum's research in action. "The collections have not lost out to architecture — they have lost to these three aspirations," he told *Nature*.

"The public will get an unprecedented insight into how the institution's painstaking taxonomic research is carried out."

The Friends of the Natural History Museum will meet with museum director Michael Dixon on 22 May to discuss the plans. "Some fairly pointed questions need to be asked," says Barlow.

Chief among these will be why the new storage facility, known as the 'cocoon' after its curving walls encased in an eight-storey glass box, will have only 3.4 kilometres of shelving for storing specimens, even though a 2001 report by the museum's trustees said that 4.6 kilometres of shelving was required.

Fitton says the museum commissioned a study at the project's inception in 1999 to evaluate the idea of storing all the collections under one roof. But with the selected design, by Danish firm C. F. Møller Architects, this was considered impractical, and the 2004 unveiling of the design to museum workers was met with anger (see *Nature* 432, 659; 2004).

Researchers' fears for the collections are exacerbated by the fact that the specimens are currently being kept at the museum's overspill facility in Wandsworth, southwest London. Barlow says this building is prone to high humidity and is below the flood level of the nearby River Thames, but museum officials say conditions are first-class.

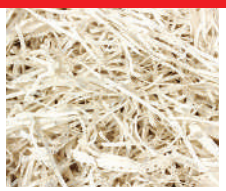
It remains unclear whether some specimens will stay in Wandsworth permanently after the new building opens, or whether space will be found elsewhere on the museum's main campus in central London. Of the 129,000 drawers in which the entomology collection currently sits, there will be room for 120,000 in the Darwin Centre, says Fitton. "We could have squeezed all the insects in but that would be silly," he says, because the collection is expected to grow by around 10% during the life of any storage facility.

Meanwhile, museum directors say the public will get an unprecedented insight into how the institution's painstaking taxonomic research is carried out. Planned attractions include an 'Explore tour', which will take visitors through selected parts of the collection and labs, and the David Attenborough studio, named after the well-known zoologist and broadcaster, which promises face-to-face interaction with museum staff. ■

Michael Hopkin

NHM

Safe storage of specimens such as those of the order Lepidoptera is a Natural History Museum concern.



ANTI-SHREDDER TACKLES SPY FILES
A computer program could reassemble notes from the East German Stasi.
www.nature.com/news

Rules tightened for aboriginal studies

Canada's main funding agency for health research has adopted ethics guidelines for studies involving the estimated one million indigenous people in the country.

Researchers and tribal leaders are optimistic that the guidelines will help to speed up the investigation of health problems among Canadian aboriginals, whose life expectancy has been estimated to be between five and eight years shorter than that of other Canadians. This is due in part to a greater susceptibility to diabetes, cardiovascular disease and mental-health problems.

W. KAEHLER/CORBIS

Improved rules could speed up research into indigenous health.

The guidelines — the first to offer a comprehensive framework for researchers working with aboriginal communities — state that scientists should focus on the collective interests of the community in planning their studies. For example, they must obtain consent for projects from the community involved, as well as from individuals. Biological samples should be considered "on loan" to the researcher, and the community should have the opportunity to review the conclusions drawn from the data.

"This is a big step forward," says Alan Bernstein, president of the Canadian Institutes of Health Research (CIHR), the main federal funding agency for health research and the author of the guidelines.

Until now, Canadian researchers working with indigenous people have relied on the guidance in a 1998 human research ethics statement issued by the Tri-Council (see *Nature* **395,** 420; 1998), a body comprising three federal research agencies, of which the CIHR is one. Ethicists have said that its comments on indigenous peoples are inadequate, and agencies have long recognized the need to beef up the guidance.

"Standards were quite uneven across the country," says Jeff Reading, scientific director of the Institute of Aboriginal Peoples' Health (IAPH), which is part of the CIHR.

The guidelines were prepared by the Ethics Office of the CIHR and the IAPH, in partnership with the aboriginal communities of First Nations, Inuit and Métis. A working group surveyed the research guidelines developed in other countries and by aboriginal groups to explore the issues raised by research involving aboriginal people.

Although the recommendations are not legally binding, health researchers and institutions funded by the CIHR will be required to follow them.

The relationship between researchers and aboriginal communities has been problematic in the past. So-called fly-in researchers would come into communities, take samples, and never be seen again. The issue was thrust into the public eye in 2000, when the Nuu-chah-nulth, a First Nations community living on Vancouver Island, British Columbia, discovered that DNA samples collected between 1982 and 1985 to study the high incidence of rheumatic disease in the community had been later used in other research projects, including an anthropological study that was not part of the original consent (R. H. Ward *et al. Proc. Natl Acad. Sci. USA* **88,** 8720–8724; 1991).

Global genetics projects such as the Human Genome Diversity Project and the Genographic Project, which both explore human origins and migrations, have also met with opposition from aboriginal groups, partly because there were no obvious benefits to the community. ■

Hannah Hoag

Canadian government pushed to protect land

Scientists from around the world have urged the Canadian government to take action to protect its northern landscapes. They claim that development is threatening one of the last great tracts of woods and wetlands on Earth.

In a 14 May letter delivered to Bob Mills, chair of the environment committee in Canada's House of Commons, 1,521 scientists from 51 countries asked the Canadian government to support the principles that were laid out in the 2003 Boreal Forest Conservation Framework.

That agreement between industry, indigenous people and conservation groups said that at least half of the 6 million square kilometres of boreal region should be off-limits to development, and the rest should be managed sustainably. The boreal region contains forest, wetland and other landscapes, and covers much of northern Canada.

"Provincial governments are left free to support unlimited development."

The campaigners say that the region is crucial to ecosystems worldwide. It serves as a breeding ground for 3 billion migratory birds, and is an important source of fresh water and the largest terrestrial carbon sink on Earth. So far, it has sequestered more than 180 billion tonnes of carbon.

One of the letter's signatories is Ken Caldeira, a climate modeller at the Carnegie Institution based at Stanford, California, who last month published a study suggesting that the total destruction of Canada's boreal forests would actually ease global warming (see *Nature*, doi:10.1038/news070409-2; 2007). Some say that press coverage of the study raised public doubts about the need to conserve the boreal region.

"It is certainly something we are struggling with," says Jeff Wells, science adviser for the International Boreal Conservation campaign in Seattle, Washington.

Canadian prime minister Stephen Harper has said that he will designate almost 100,000 square kilometres of land in the Northwest Territories as protected, but those deals have yet to be finalized. That leaves provincial governments free to support unlimited development, says David Schindler, a boreal ecologist at the University of Alberta in Edmonton. ■

Erika Check

ON THE RECORD

"I used to watch *Doctor Who* and *Star Trek*, but they went [politically correct] — making women commanders, that sort of thing. I stopped watching."

British television personality Patrick Moore, who has presented a show on astronomy since 1957, alienates a large proportion of his audience.

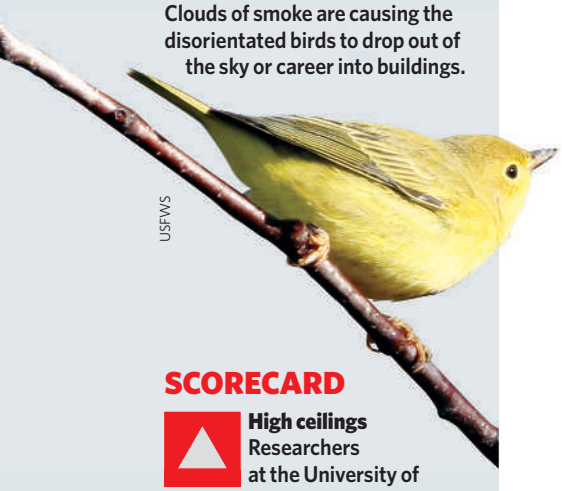
"I believe in evolution. But I also believe, when I hike the Grand Canyon and see it at sunset, that the hand of God is there also."

US presidential hopeful John McCain (Republican) tries not to alienate anyone at a recent debate. Three of his rivals for the nomination declared that they don't believe in evolution.

ZOO NEWS

Crash and burn

Wildlife workers in Florida say that the wildfires raging across the state are killing hundreds of yellow warblers (pictured). Clouds of smoke are causing the disorientated birds to drop out of the sky or career into buildings.



USFWS

SCORECARD

High ceilings

Researchers at the University of Minnesota claim that rooms with lofty ceilings encourage "freer, more abstract thinking", whereas more poky spaces activate more "confined concepts".

Texan skies

Plans for a 'body farm' to help forensic experts study decomposition have been shelved amid fears that the Texas State University facility would attract buzzards and endanger nearby planes.

Sources: ABC, BBC, Livescience.com, washingtonpost.com, Associated Press

SIDELINES

Applicants challenge male order at Howard Hughes

The Howard Hughes Medical Institute (HHMI) is one of the wealthiest philanthropic organizations supporting biomedical research, but Dora Angelaki didn't want to become one of its investigators just for the money.

With funding from the National Institutes of Health (NIH) and NASA, her neurobiology lab at Washington University in St Louis, Missouri, is hardly starved for cash. What she wanted was to embark on riskier projects that would be difficult to sell to conservative, federal research agencies. So when her colleagues encouraged her to enter the HHMI competition, she was eager to give it a go.

Unfortunately, as Angelaki soon discovered, the competition is open only to academic researchers who have been running a lab for between four and ten years. Angelaki has been at it since November 1993. And even with the two extra years that the HHMI allows to compensate for the fact that she had two children during that time, Angelaki missed the window by 18 months.

"The four-to-ten-year window works very strongly against women."

Her exclusion highlights what critics claim is a persistent problem for women scientists trying to become an HHMI investigator. Some argue that the institute's nominating practices and restrictive time limits hinder such career development — issues that the HHMI itself has recognized and taken steps to correct. But questions persist as to whether these steps go far enough.

"They've taken a huge step forward," says Ben Barres, a neurobiologist at Stanford University in California. "But they still need to diversify their selection committees. And I think this four-to-ten-year window works very strongly against women."

As a result of changes put forward in 2005 by the institute's president, Thomas Cech, the current competition is the first in which individuals can directly apply to the HHMI, rather than having to be nominated by a university. This is intended to make the process more open to women, minority groups and young scientists.

Nevertheless, some researchers would like the HHMI to go further. Barres says it should extend its window of eligibility from 10 to 16 years, to lessen the disadvantage faced by women who have had children soon after becoming assistant professors. But HHMI spokeswoman Avice Meehan says that the 10-year limit is important to enable the institute to maintain a pool of early-career investigators. And, she adds, applicants for the investigator competition are

Time runs short for HapMap

Geneticists at the genomics meeting in Cold Spring Harbor, New York, last week celebrated the success of the International HapMap project — and predicted its eventual demise.

The project, unveiled in 2005, is a database of markers of genetic diversity called single nucleotide polymorphisms, or SNPs. By comparing HapMap data against SNPs from people with specific diseases, researchers hope to pinpoint the genetic glitches that underlie those diseases.

Many studies have now been published that have used data from the HapMap project, which aims to cut down the cost and effort involved in finding the genetic errors that lead to common disorders such as heart problems, mental illness and diabetes.

At the meeting, Peter Donnelly of the University of Oxford, UK, who is one of the investigators in the Wellcome Trust Case Control Consortium, presented findings from a major study the group conducted of seven common diseases in 16,179 participants. The study, expected to be published shortly, is perhaps the most ambitious trawl yet for common disease genes. And it did turn up possible genetic causes for some of the diseases, Donnelly said. Together with results from other 'genome-wide association studies', Donnelly's report had researchers toasting their successes, and hoping for more.

"For ten years, we have been predicting that genome-wide association studies will work, and it is clear now that they do," says Ewan Birney, who runs the gene bank at the

Critics say it remains difficult for women to become investigators with the Howard Hughes Medical Institute.

P. FETTERS/HHMI

eligible for extensions on the basis of childcare, parent care or military service.

At present, 62 out of 299 HHMI investigators are female. That's not much different from the balance in the pool of faculty members from which the HHMI draws its candidates: at US research universities, women accounted for 30% of assistant and associate professors and just less than 20% of full professors in the life sciences in 2003, according to the National Academies. That means it will be hard to dramatically change the percentage at the HHMI, concedes Nancy Hopkins, a developmental biologist at the Massachusetts Institute of Technology in Cambridge.

But Hopkins thinks it would be a good idea if the HHMI put more women on its powerful executive boards. "It's these powerful organizations and the people at the top that are really visible and send a strong message," she says. Two of the eight members of the medical advisory board and four of the nineteen members of the scientific review board are female. Among the ten HHMI executive officers, Meehan is the only woman. "That is a male organization," Hopkins says. "It just is."

Cech has already pledged to increase the percentage of women on the review panels that select investigators. These panels have been between 22% and 30% female, but the panel for the 2008 competition has not yet been selected, says Meehan.

The community will be watching closely to see whether the present changes have the intended effect, but for Angelaki, they are too late. When she got notice that she was not eligible to become an HHMI investigator, she sent an e-mail back. "Thank you for your response," she wrote. "It's too bad for both of us." ■

Heidi Ledford

European Molecular Biology Laboratory in Cambridge, UK. "What this meeting says to me is that we have got to do this in every disease that matters to us, including diseases of the developing world."

But another message from the meeting was that tools not available at the dawn of the HapMap could soon make the project obsolete. Scientists described some of the first sequencing projects that use new technologies developed to decode DNA faster and more cheaply than ever before.

The new technologies will allow researchers to compare many more types of variation than just SNPs, and in greater numbers of people. The fast sequencers available now — made by 454 Life Sciences in Connecticut, and Solexa in San Diego, California, both of which were acquired this year by larger corporations

"The HapMap was a substitute until we could afford to do whole genomes."

— enable scientists to read the genomes of small organisms in weeks, and of larger organisms in months. For instance, 454 is expected to announce shortly that its technology was able to start and complete the sequencing of the genome of geneticist James Watson between January and March.

Over the next year, new rapid sequencers from Applied Biosystems in Foster City, California, and Helicos Biosciences in Cambridge, Massachusetts, are expected to make their debut. Once these quick sequencers become routine, it will obviate the need for tools such as the HapMap, scientists say. "The HapMap was a substitute until we could afford to do whole genomes," says plant genomicist Magnus Nordborg of the University of Southern California in Los Angeles. ■

Erika Check

Charity announces deal for cheap AIDS drugs

Two steps taken last week may help keep the price of AIDS drugs down in developing countries.

In one development, the charity run by former US President Bill Clinton announced that it has negotiated price reductions of up to 67% for 16 AIDS drugs in 66 developing countries. The New York-based William J. Clinton Foundation says the cut-price medicines will be supplied by two Indian manufacturers of generic drugs, Cipla and Matrix Laboratories, and paid for by UNITAID, an international drug-purchase facility established in 2006 by France, Brazil, Britain, Norway and Chile.

The 8 May announcement came shortly after Brazil said it would break a patent on efavirenz, an AIDS drug made by Merck. Under some rarely used World Trade Organization rules, a country may, in a public-health emergency, sign a compulsory licence that allows it to make or buy generic versions of patented drugs. Merck, which had offered the Brazilian government a 30% discount on the pills, said it was "profoundly disappointed" with the decision.

Space satellite brings continents together

Africa and Asia joined forces on 14 May as the Chinese Long March 3-B rocket carried a telecommunications satellite into space for Nigeria.

China's state news agency says it is the first time that a foreign buyer has purchased both a satellite and its launching service from the country. About 30 other launches are planned, including one for Venezuela next year, the agency says.

NIGCOMSAT-1 now joins Nigeria's Earth observation satellite NIGSAT-1, which was launched from Russia in 2003.

LI GANG, XINHUA/AP

Long March to space: China this week launched a Nigerian telecommunications satellite into space.

Massachusetts proposes $1 billion for life sciences

S. SENNE/AP

Deval Patrick: planning a boost for biotech.

Massachusetts governor Deval Patrick has proposed a US$1-billion life-sciences initiative intended to sustain the state's prominent biotechnology industry in the face of stagnant federal funding and increasing competition from other states, notably California.

The plan, unveiled on 8 May at the BIO biotechnology convention in Boston, calls for funding over the course of 10 years and includes: tax incentives for biotech companies; a competitive grant programme for biomedical researchers; $38 million in funding for RNA-interference research; and $66 million to set up a stem-cell repository.

Funding for the proposal hinges upon legislative approval, but the heads of both houses of the Massachusetts legislature have said they will support the initiative.

Nuclear-weapons lab under new management

Security concerns at the Los Alamos National Laboratory in New Mexico have had the entire US nuclear-weapons complex making big changes. Last week, for the first time in more than half a century, the Lawrence Livermore National Laboratory in California got a new manager.

The new boss looks much like the old one, though: it is a consortium led by the University of California, which has managed the lab since it was founded in 1952, and the Bechtel Corporation.

The group — essentially the same team that won the right to manage Los Alamos in December 2005 — narrowly beat a rival bid from an industrial group led by the defence giant Northrop Grumman. The University of California/Bechtel proposal to manage the lab — which has a yearly budget of US$1.6 billion — was lower in cost and scored higher, in part because of the university's scientific expertise, according to Tyler Przybylek, the Department of Energy official who chose the consortium.

Under the seven-year agreement, the consortium will receive up to $45.5 million per year to oversee the lab, which was recently chosen to develop the reliable replacement warhead, a new class of nuclear warhead that will not require testing (see *Nature* 446, 121; 2007).

Publishing prize

Annette Thomas, the managing director of Nature Publishing Group, was last week awarded the 2007 Kim Scott Walwyn Prize for women in publishing in the United Kingdom. Annette was a biology editor at *Nature* and the founding editor of *Nature Cell Biology*, before moving into publishing with the launch of the *Nature Reviews* journals.

Medical Research Council emphasizes clinical work

Britain's Medical Research Council (MRC) is set to shift its focus towards 'translational' research, which has a greater emphasis on producing clinical benefits for patients and is currently very much in vogue at the US National Institutes of Health (see *Nature* 441, 17–19; 2006). The change comes two months after the British government announced a 2.7% rise in the MRC's science budget for next year.

On 10 May, MRC leaders announced that it plans to improve links with the pharmaceutical industry and potentially increase the number of drugs it brings to market itself. But outgoing chief executive Colin Blakemore assured basic researchers that they will not lose out. "I don't think there's cause for concern," he said. "This is not about withdrawing from areas we have previously funded."

Iranian researchers to rejoin chemical society

The American Chemical Society has reinstated 14 of the 36 Iranian researchers who had their memberships cancelled earlier this year (see *Nature* 446, 597; 2007).

The decision was made after consultation with lawyers and the US Department of the Treasury, which currently has a near-total embargo on dealings with Iran. The reinstated members will be eligible for all benefits except career-development services and discounts for meetings. The remaining 22 members will be offered an opportunity to renew their memberships, which have lapsed since the cancellations were made.

Correction

Our News Feature 'Is this what it takes to save the world?' (*Nature* **447,** 132–136; 2007) incorrectly identified Ralph Cicerone as a co-winner of the Nobel prize shared by Paul J. Crutzen, Mario J. Molina and F. Sherwood Rowland.

BUSINESS

Digging deep

Long dismissed as too expensive or impractical, mining the sea floor for metals is gaining a new foothold. **Mark Schrope** reports on two companies hoping to take the plunge.

When the riches of the land run out, look to those of the sea. That's how many have thought for years, although mining the deep sea for metals always seemed a prospect of the distant future. Now, two companies are making a bid to turn dreams into profits by mining massive deposits of copper, zinc, gold and silver on the sea bed.

These 'seafloor massive sulphides' the companies are targeting lie at extinct hydrothermal vents, and can be more than an order of magnitude richer in the metals than currently active mines on land. But concerns have been raised about the environmental effects of pressing forward with mining. The companies involved have funded the most extensive studies ever of extinct vent areas, and some scientists are cautiously optimistic that the activity will have only a minimal effect on the surrounding area.

Leading the new wave is Nautilus Minerals, founded around 1997 and based in Vancouver, Canada. The company has raised US$300 million in funding and its stock price has more than doubled since it became a public company in 2006. Other mining companies, including Anglo American, Teck Cominco, Epion and Barrick Gold, have all signed on as stockholders. Nautilus's president, David Heydon, dreams that sulphide mining on the sea floor will become as big as oil drilling is there. The company hopes to start mining operations by the end of 2009 in the Bismarck Sea on the north side of Papua New Guinea, pending approval of an environmental assessment to be submitted later this year.

Branching out

Meanwhile, a smaller, London-based company called Neptune Minerals, which started in 1999 and went public in 2005, has raised roughly $18 million and taken a different approach. Rather than focus on one area, it continues to actively explore a number of regions, and has applied for or received exploration leases in Japan, New Zealand, Micronesia and elsewhere.

Both companies have so far stuck to the complex permitting process for mining in countries' territorial waters. If they were to extend their activity into international waters, the mining would be governed by the strict guidelines established under the International Seabed Authority in Kingston, Jamaica.

The new push has worried environmentalists, who had thought that deep-sea mining for metals was a dream of the future. "I was kind of hoping it would be economically unfeasible for a long time," says Janet Voight, a deep-sea biologist with the Field Museum of Chicago, Illinois, who has studied extinct vents in the North Pacific Ocean.

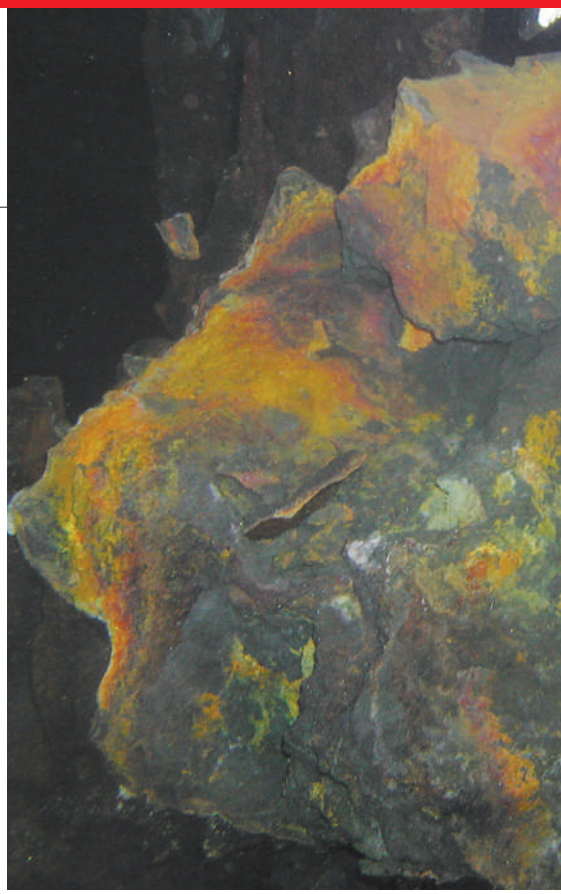
Strong markets

For decades starting in the 1960s, deep-sea mining for metals was all the rage — over manganese nodules on deep-sea abyssal plains as potentially rich sources of several metals. But those prospects never panned out into full-scale, economically feasible mining, in part because of declining metal prices through the 1980s and 1990s. Today, experts say that the market is much stronger for materials recovered from sulphide mining — especially because of developing countries' demand for copper for electrical wire. "I don't think there's any doubt that the markets are going to stay strong for the metals," says Caitlyn Antrim, executive director of the Rule of Law Committee for the Oceans, based in Washington DC.

> **"I don't think there's any doubt that the markets are going to stay strong for the metals."**
> — Caitlyn Antrim

Getting to the metals is a challenge. Seafloor massive sulphides are the product of hydrothermal vents, whose magma-heated waters are rich in minerals that precipitate out on the sea floor once they meet the near-freezing waters of the deep sea. The deposits are thus exposed in vast expanses, unlike buried ore in land mines, and easier to collect if they can be reached.

The Nautilus mining plan calls for tethered remotely operated vehicles with cutting heads, working at first in about 1,600 metres of water, to mulch the material into baseball-sized chunks. These will then be carried via suction pipes to hoppers on a ship. The cold water will be filtered and pumped back below 500 metres, while the ore is barged to shore for processing and shipment to smelters in other countries. Nautilus plans to mine a region just 200 metres by 200 metres and 15 metres deep in a year. Heydon says that its yield will match that generated from 40 times more material obtained from land-based mines because the sulphides are richer and more accessible. Neptune is still considering technologies for operations that it hopes to begin in 2010.



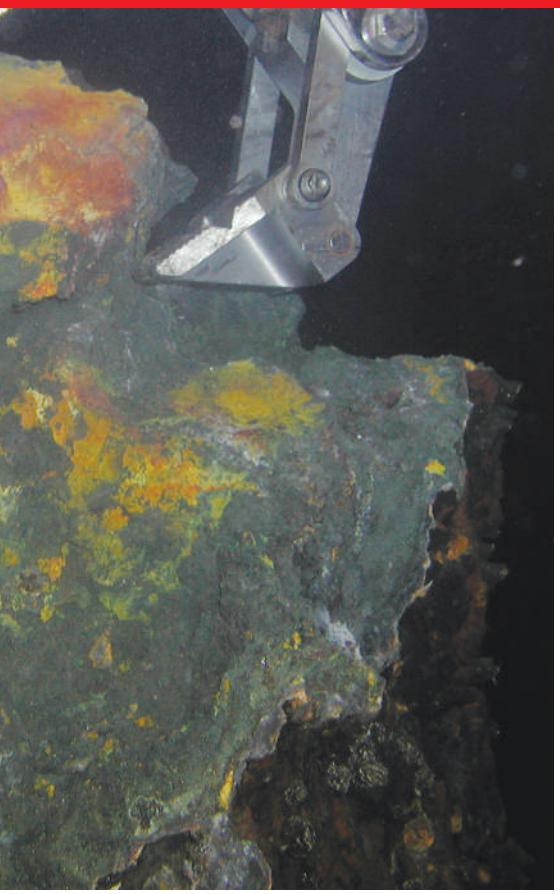
Down and out: remotely operated vehicles can extract metals such as copper from the sea floor.

One of the side effects of deep-sea mining for metals has been a rush in studies of extinct hydrothermal vents, funded almost exclusively by the companies. Nautilus, for instance, has sponsored several research cruises to Papua New Guinea and has supported a cruise led by the Woods Hole Oceanographic Institution in Massachusetts.

Safety first

Among other concerns, scientists have questioned whether the mining would harm endemic or rare organisms, disrupt important connections between systems or create plumes that affect benthic filter-feeders and midwater organisms. Heydon thinks that the bulk of the data needed to show the work can proceed safely is already in hand. Others are not yet convinced.

Cindy Van Dover, director of the Duke University Marine Laboratory in Beaufort, North Carolina, is one of the biologists contracted by Nautilus to study organisms at the extinct vents. Based on what she calls simple studies with a remotely operated vehicle, corals and barnacles seem to be the most conspicuous residents. Although her work has not identified any major red flags over the mining, Van Dover is not yet comfortable with the thought of mining commencing. "We have got to understand



NAUTILUS MINERALS

so much more about these systems," she says. Both Van Dover and Voight are concerned about the impacts if Nautilus is successful and mining activity spreads substantially.

As one way to address concerns, Nautilus plans to set aside parts of the fields in its Greece-sized lease area as preserves, and one goal of the environmental assessment work has been to compare the proposed mining and preserve sites to ensure that they have similar species composition. Martin Angel, a midwater biologist from the Southampton National Oceanography Centre, UK, has some reservations, but says that his general impression "is that the impact will probably be minimal."

"My first impression is that Nautilus is trying to do the right thing," adds Paul Lokani, who heads the Nature Conservancy's Papua New Guinea office and voiced concerns about the plans at a stakeholders' meeting in Port Moresby in March.

Beyond biology, Heydon says that taking minerals from ore in the sea floor will be more environmentally friendly than land-based mining. Needing to remove less material for the same amount of metal will dramatically reduce energy needs, for instance. Rod Fujita, with New York-based Environmental Defense, says that such broader benefits are worth considering. However, he cautions that they should be weighed against the relative benefits of land mining, such as greater accountability due to visibility.

Van Dover says that the scientific community has little time but a rare opportunity at the outset of the work. "Figuring out how it gets done right, if it gets done," she says, "is very important." ■

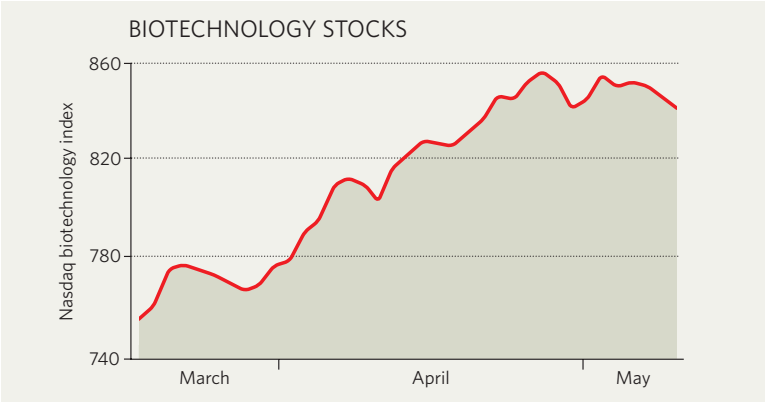
IN BRIEF

GOING GREEN General Motors in Detroit, Michigan, has signed onto a group of other businesses calling for a 60–80% reduction in greenhouse-gas emissions by 2050. The firm is the first carmaker to join the US Climate Action Partnership, which already includes companies such as DuPont and Duke Energy and has called for legislation to regulate emissions in the United States. PepsiCo and Shell Oil were among the 14 new members who joined the group last week. Environmentalists praised the move but said that only formal legislation could enforce such caps.

FINDING ITS WAY Europe's beleaguered Galileo satellite navigation system sustained another blow on 10 May, when the eight-company consortium charged with building the system had the project taken away from it. Instead, the European Union plans to take over the €3.6-billion (US$4.9 billion) project. The satellites, a competitor to the US Global Positioning System, are meant to be operational by 2012 — a deadline that the European Union says it intends to honour even if it means taking on additional costs.

DRUG DOINGS The US Senate voted last week 93–1 in favour of a bill that significantly boosts the power of the US Food and Drug Administration (FDA) to police the safety of drugs once they are on the market. The bill was modified from an earlier, more stringent version (see *Nature* 446, 844–845; 2007) — for instance, the FDA will not have the option of enacting a two-year moratorium on direct-to-consumer advertising of new drugs. The legislation now moves to the House of Representatives, where a vote is expected this summer. Senators Edward Kennedy (Democrat, Massachusetts) and Mike Enzi (Republican, Wyoming) led the charge for the bill.

MARKET WATCH



This week, Wood Mackenzie, a research and consulting firm based in Edinburgh, UK, reviews recent trends in biotechnology stocks.

After a sharp drop in late February caused by a market correction that saw indices plunge in all fields, by mid-March the Nasdaq biotechnology index resumed the overall rise that has been its general trend since July 2006.

In the past two months many companies have published their annual reports from last year, as well as first-quarter results for 2007, and a generally strong showing has underwritten the rise in the index. Two winners were Californian heavyweight Gilead, driven by strong sales of its HIV drug portfolio, and New Jersey's Celgene, fuelled by strong sales of its blood-cancer drug, Revlimid, who both saw their share prices rise in value by almost a quarter.

But the big news is the acquisition of one of biotechnology's key players, Maryland's MedImmune, by UK pharma giant AstraZeneca in late April for US$15.6 billion (see *Nature* **447,** 131; 2007). AstraZeneca wanted biological manufacturing expertise and infrastructure, whereas MedImmune has struggled with a weak pipeline of late-stage drug candidates. MedImmune's share price, already rising as a result of strong sales in the first quarter, gained 18% on news of the offer — which must still obtain regulatory approval. Meanwhile, Californian diagnostics company Biosite is the subject of a bidding war between California's Beckman Coulter and Massachusetts' Inverness Medical Innovations, causing Biosite's share price to rise by almost four-fifths over the past 2 months.

Because the market for public offerings in biotechnology remains tough, mergers and acquisitions are increasingly the life blood of the industry — and the preferred exit for investors. ■

NATURE|Vol 447|17 May 2007

STATE OF THE DONATION

Wealthy philanthropists and private foundations are supporting biomedical research on a grand scale. **Meredith Wadman** asks what they get for their money.

B. MELLOR

Scientists arriving at the Stowers Institute for Medical Research in Kansas City, Missouri, might think they've chanced on Xanadu. Limestone floors and fine furniture seduce the eye, and from the expansive gardens comes the soothing sound of fountains. Inside this research palace, funded by a $2-billion endowment from local mutual-fund magnates Jim and Virginia Stowers, scientists pursue research on fundamental cell biology. Dozens are flush with US$1 million funding a year, and their work is often destined for stellar publications.

The Stowers are not alone in their generosity; philanthropic foundations have long had an important role in biomedicine, from the birth of the Carnegie Institution in Washington DC in 1902 to the biomedical activism of the Rockefeller Foundation in the 1930s and 1940s. But during the past decade, philanthropists — and the foundations that they establish to distribute their money — have begun funding biomedical science on a particularly striking scale.

Dominating this landscape is the Bill & Melinda Gates Foundation with its plans to nurse the globe to better health, boosted last year by billionaire financier Warren Buffet (see page 254). But there are other donor organizations, many of which fund more basic biological research, ranging from the United States' mighty Howard Hughes Medical Institute (HHMI) to Britain's Wellcome Trust, the world's largest charity exclusively devoted to biomedicine (see page 251). Their cash is being lapped up by researchers parched by flat funding from the US National Institutes of Health (NIH) and many other sources.

> **"Foundations may make decisions about public policy without any public discussion or political process."**
> — Pablo Eisenberg

These new givers — the gigaphilanthropists — are perceived to be making an impact on the research landscape that is much greater than the sum of their dollars. "The effect of the private foundations is not reflected in the total funding they supply. They have disproportionate influence," says Hamilton Moses of the Alerion Institute, a Virginia-based think-tank that focuses on innovation in biomedical research. They can, and do, take financial and scientific risks unthinkable with tax-payers' dollars. They fill gaps left by government and industry, dictate exactly what their money is spent on and act quickly compared with the sometimes glacial pace of government agencies. But although those running the organizations are sure that private money buys more and better science than public money, there is little concrete evidence they are right.

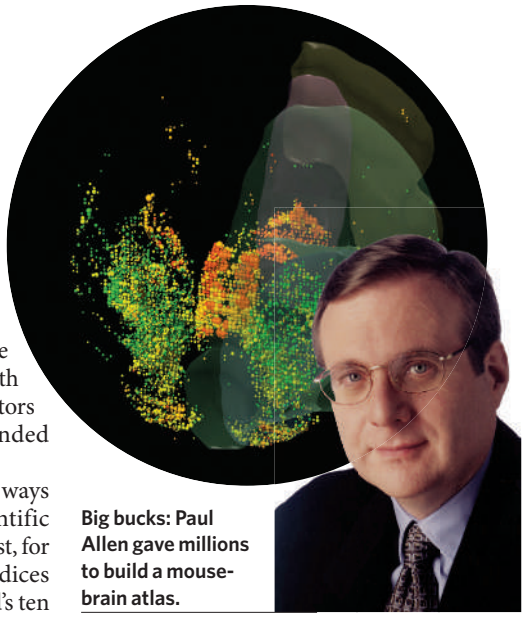
The new wealth also comes with strings attached: some funders keep a businesslike control over the direction of the research they pay for and demand a level of accountability that can make researchers uncomfortable (see page 252). Some observers worry about the growing power wielded by the gigaphilanthropists over the research agenda if, as is predicted, charitable giving reaches new heights in the future. They are concerned that too many important decisions with an impact on biomedicine will be made in the boardrooms of foundations with little scientific expertise — and no public input or accountability. "You may have foundations with assets larger than almost 70% of the world's nations making decisions about public policy and public priorities without any public discussion or political process," says Pablo Eisenberg, a senior fellow and philanthropy-watcher at the Georgetown Public Policy Institute in Washington DC.

Value for money?

By all accounts, the amount of money from non-profit foundations and philanthropists is growing strikingly. In Germany, for example, the Frankfurt-based Hertie Foundation has spent more than €90 million (US$122 million) on nurturing neuroscience since 2000, compared with a total of €30 million in the previous quarter century.

In Britain, charities fuel more than half of the biomedical research enterprise, led by the Wellcome Trust, which served up £484 million (US$960 million) in research funding last year compared with £270 million just over a decade ago. In the United States, a wave of philanthropic giving over the past decade has been fuelled by a buoyant stock market and a generation of ageing, affluent baby boomers. Investing in biomedicine allows them to do something that might support the health of their children, and gains them significant tax breaks.

In 2005, a study led by Moses showed that private, non-corporate support for US biomedical



VULCAN/ALLEN INST. BRAIN SCI.

Big bucks: Paul Allen gave millions to build a mouse-brain atlas.

research leapt 36% to $2.5 billion between 1994 and 2003 (H. Moses *et al. J. Am. Med. Assoc.* **294,** 1333–1342; 2005). Today, "$5 billion is probably an undercount", Moses says, when one includes philanthropic funding in all its varieties. This still accounts for no more than 5% of the roughly $100 billion spent annually in the United States on biomedical research (the biotechnology and pharmaceutical industries account for about 60% of this total and the government, led by the $29-billion budget of the NIH, for most of the rest). But even so, the massive donations and influence of the US foundations, along with the Wellcome Trust, bear examination.

Linheng Li was one of the first scientists through the door after the Stowers Institute opened in late 2000. Li was intent on ending a 25-year quest by stem-cell scientists to find the 'niche' in bone marrow that harbours blood-forming stem cells. In 2003, Li delivered, with a paper in *Nature* describing the cells' physical and biochemical environment in mice (J. Zhang *et al. Nature* **425,** 836–841; 2003), a discovery that had the potential to help researchers grow stem cells outside the body.

Scores of such anecdotes suggest that the munificent backing of a gigaphilanthropist generates more, and more influential, scientific results. But they are just that: anecdotes. There are few if any studies rigorously comparing the productivity of an HHMI investigator, for example, against that of an NIH-funded colleague down the hall.

Such assessments are difficult to make, partly because there are few researchers who rely solely on a single source of funds, philanthropic or public, and so could be sensibly compared. There is also an 'apples and pears' problem, adds Mark Walport, the Wellcome Trust's director. The trust focuses much of its support on young scientists and building research capacity in the developing world. "It would not be meaningful," he says, to try to compare the scientific outputs of these programmes with those from senior, independent investigators supported by Britain's government-funded research councils.

The gigaphilanthropists do find other ways to gauge whether they are getting scientific value for their money. The Wellcome Trust, for instance, has combed global citation indices to establish that it funds five of the world's ten most-cited malaria researchers, and four out of the top ten in the cognitive sciences. The Ellison Medical Foundation, based in Bethesda, Maryland, funds ageing research. It is building an electronic archive of every application it receives and it plans — many years hence — to use citations and other measures to analyse the impact of researchers it backed and those it turned down. Meanwhile, executive director Richard Sprott says he attends as many national and international meetings as he can. "I listen to who's doing the cutting-edge, exciting stuff. If it's our people, I think we're doing okay."

"I listen to who's doing the cutting-edge, exciting stuff. If it's our people, I think we're doing okay."
— Richard Sprott

Tom Cech, the Nobel-prizewinning chemist who has directed the HHMI since 2000, says that the number of plaudits won by its scientists "prove that our investigators are far more successful than average". Between 1994 and 2007, 89 HHMI investigators were elected to the National Academy of Sciences (NAS) and seven won Nobel prizes. The HHMI has calculated that its investigators were more than ten times more likely to be elected to the NAS than US biologists funded by the NIH, and over 16 times more likely to win a Nobel prize in chemistry or medicine. "What we can't rigorously prove is whether they are more productive because of our support, or simply because we know how to choose winners," Cech says.

D. IPOCK

The Stower's Institute provides generous funding but expects star results in return.

Taking risks

Not every foundation does know how to choose winners, points out Eisenberg. "There are foundation officers who are sharp and knowledgeable, and those who are not," he says. Without more rigorous comparisons, some observers question whether, dollar for dollar, philanthropic donations guarantee more good science than government or industry does. "I don't think the data overall would hold up," says Mary Woolley, president of Research!America, a health-research advocacy group in Alexandria, Virginia. "Plenty of people who have received the Nobel prize were funded by the NIH."

Philanthropic organizations can certainly put great pressure on their grant recipients to ensure they deliver. At the Stowers Institute, the senior scientists are appraised after five years' generous funding, when the institute enlists leaders in a researcher's speciality to evaluate their performance. "The only question we ask is whether the leaders in the field can say definitively, specifically and discretely: 'This is what this person has done at Stowers that has changed how people think about the field'," says Stowers' president Bill Neaves. Since the institute opened its doors, eight of nine senior scientists who have been evaluated have passed that test.

Private foundations have a flexibility and agility with their spending that industry and government agencies do not. They are not answerable to shareholders or venture capitalists; nor do they labour under the political and public scrutiny experienced by the NIH and

P. FETTERS/HHMI

Taking a risk: Janelia Farm represents a leap in the dark both for researchers and the HHMI.

other spenders of public money. "When I was at the NIH, we had to ask ourselves a question when contemplating every award: 'Can I live with it on the front page of *The Washington Post*?'" says Sprott, a former director of the National Institute on Aging. That kind of thought process, he says, "tends to make the NIH very conservative".

The NIH, under director Elias Zerhouni, has launched an ambitious effort to battle this conservatism. Since 2004, for instance, Zerhouni has awarded 35 Pioneer awards to individual investigators for high-risk research, delivering around $2.5 million to each over the course of five years. But for philanthropists, risk-taking is often the rule rather than the exception. "When you are a very small slice of a large pie you not only have the opportunity, but I would say the responsibility, to do something out on the edge," says Cech.

Brain storm

Lately, Cech's organization has been pushing the edge at Janelia Farm, the $500-million research complex near Washington DC. Since September a cadre of top-tier scientists has been set loose here to try to unravel how information is processed by neuronal circuits — a departure for the HHMI, which until now has supported researchers in their home institutions.

At Janelia Farm, researchers work in small groups without the benefit of tenure or outside funding — but with a freedom from the hassles of grant-writing, teaching and administrative duties that is almost unique in US science. Director Gerry Rubin says that the best science comes from making such an expensive but risky investment, even if nine out of ten projects fail. "We are venture capitalists here," he notes.

The gigaphilanthropists can also move fast. When the US Postal Service was beset by anthrax attacks in the autumn of 2001, researchers quickly realized that they needed to know the background level of anthrax in post offices around the country. "It would have taken the NIH two or three years to solicit and award a grant answering that question," says Sprott. "We could pick up the phone and call the world's top anthrax expert and ask him to design and carry out the needed study. We had an answer within six months." Ellison's investigator found 15 different strains of anthrax of the kind that infect cattle, sheep and horses; none was the strain being sent through the mail.

Generally, the new philanthropists are not the type to write a cheque and walk away. They are determined to identify and fill key gaps in public funding — and to make sure the work gets done. Take Paul Allen, the Seattle billionaire who co-founded Microsoft with Bill Gates. In 2001, Allen summoned the best and brightest minds he could find in genomics, neuroscience and psychology and asked them what could and should be done to change the field of brain science. In 2003, he and his sister, Jody Allen Patton, signed a cheque for $100 million to launch the Allen Institute for Brain Science in Seattle*. Last year, researchers there unveiled the Allen Brain Atlas, a three-dimensional map showing where thousands of genes are active in the mouse brain. Some 800 scientists are using the atlas daily, according to the institute.

Like the Allens, foundations "all want missions accomplished", says Donald Brown, president of the Life Sciences Research Foundation, a non-profit organization based in Baltimore, Maryland, that solicits grants from foundations and industry to support postdocs.

WELLCOME PHOTO LIB.

The Wellcome Institute's Mark Walport has hundreds of million of pounds to spend.

Targeted philanthropic spending gets big, expensive projects done fast, but it also draws criticism. Because US charitable foundations are required by law to spend 5% of their assets each year, money can be thrown at projects too quickly for some people's tastes. Gigaphilanthropists can choose to fund research at the whim of their benefactors and the advisers they choose, very different to the extensive consultation with the scientific community that occurs before government money is committed to a big project. In the case of the Allen Brain Atlas, some researchers grumble that the money would have had a far greater impact on neuroscience if it had been spread among a group of top investigators.

More recently, some foundations' alacrity has been taxed as they find themselves scrambling to compensate for the funding freeze at the NIH and keep individual investigators afloat. Some are concerned that this constrains their flexibility — and that they risk being taken for granted. "I worry that we are allowing the government to say: 'We don't have to pay the bill. Private philanthropy will step in,'" says Sprott. "That's not what we want to do."

As Paul Schervish sees it, the current flow of philanthropic money may look a mere trickle by the time today's postdocs retire. In a widely cited 1999 report, Schervish, director of the Center on Wealth and Philanthropy at Boston College, and his colleague John Havens predicted that by 2052 at least $6 trillion in wealth would be transferred from the estates of older Americans to charity — some $100 billion more per year than today. That would amount to a significant chunk of cash for biomedicine if, as is the case today, roughly 20% of that money goes to health. "We are going to see substantial foundation growth," says Schervish.

If his prediction pans out, many more researchers will find themselves relying on — and answerable to — the gigaphilanthropists. At the Stowers Institute, the founders have already announced plans to add 56,000 square metres of facilities and 600 people every decade in perpetuity. But in the rich biomedical landscape of the future, it may be just one of many palatial shrines to research. ■

Meredith Wadman is a biomedical reporter for *Nature* in Washington DC.

For more on philanthropy, see our online special at www.nature.com/news/specials/philanthropy. See also Editorial, page 231.

Nature has a commercial collaboration with the Allen Brain Institute in the Neuroscience Gateway.

BILL & MELINDA GATES FOUNDATION (US)

2006 spend: US$908 million

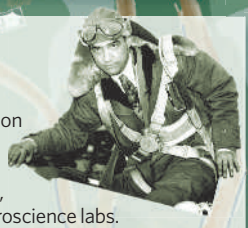
The largest charitable foundation in the United States grew from Microsoft profits and some of Warren Buffett's investing fortune. The figure above comprises both research and non-research spending in its Global Health Program.

WELLCOME TRUST (UK)

2006 spend: £484 million (US$960 million)

Long seen as the saviour of British science, the world's biggest biomedical charity also focuses on building research capacity in areas such as Africa and southeast Asia.

HOWARD HUGHES MEDICAL INSTITUTE (US)



2006 spend: US$694 million

Renowned for the generous grants it bestows on the young and gifted, the HHMI is the second richest biomedical charity in the world. Its new Janelia Farm campus boasts games tables, StairMasters and a pub interspersed with neuroscience labs.

AFM (France)

2006 spend: €61 million (US$82 million)

The French Muscular Dystrophy Association (AFM) is France's main private funder of research in genetic disorders. It funded the first maps of markers across the human genome.

HIGH Q AND CHDI (US)

2006 spend: US$50 million

These sister foundations have borrowed tactics from biotechnology companies in an attempt to fast-track the development of drugs against the fatal, inherited brain condition called Huntington's disease.

STOWERS INSTITUTE FOR MEDICAL RESEARCH (US)

2006 spend: US$73 million

Cancer survivors James and Virginia Stowers have thrown their earnings from the mutual-funds industry at biomedical causes. In addition to founding the $2-billion Stowers Institute, they spent more than $15 million last year on a political campaign to advance stem-cell research in Missouri.

ALLEN INSTITUTE FOR BRAIN SCIENCE (US)

2006 spend: US$13 million

Microsoft co-founder Paul Allen is fascinated by brain circuitry. Last year, the institute he founded published the Allen Brain Atlas, an online map of gene expression in the mouse brain (left).

ELLISON MEDICAL FOUNDATION (US)

2006 spend: US$36 million

Larry Ellison, chief executive of software giant Oracle, spent two weeks in the lab of Nobel laureate Joshua Lederberg in 1994. A portion of his fortune now funds researchers who study ageing.

CLIVE AND VERA RAMACIOTTI FOUNDATION (Australia)

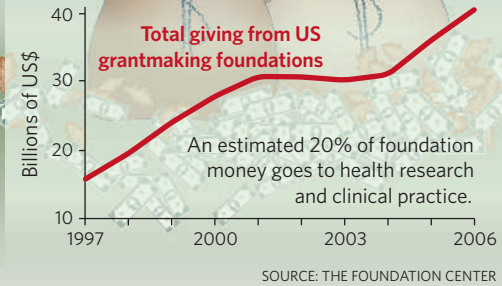
2006 spend: Aus$1.3 million (US$1 million)

In 1970 Vera Ramaciotti sold a theatre left to her and her brother Clive, and used the proceeds of Aus$6.7 million to set up what has become Australia's largest charitable foundation dedicated to biomedicine.

THE VOLKSWAGEN FOUNDATION (Germany)

2006 spend: €5.6 million (US$7.6 million)

Founded with proceeds from a Volkswagen factory, this is one of the ten richest foundations in the European Union. The figure here is spending on biomedical research. It supports interdisciplinary research and work excluded from funding by other sources.



SOURCE: THE FOUNDATION CENTER

The money tree

Donations from philanthropists and private foundations are increasingly finding their way into biomedical research. **Lucy Odling-Smee** takes a look at some of the richest and most influential funders.

ILLUSTRATION: N. SPENCER. PHOTOS: BILL & MELINDA GATES FOUNDATION; WELLCOME PHOTO LIB.; BETTMANN/CORBIS; L. PSIHOYOS/CORBIS; ALLEN INST. BRAIN SCI.; VOLKSWAGEN OF AMERICA/AP

LOVE OR MONEY

Biomedical scientists want funding; private foundations want cures. **Erika Check** hears the joys and tensions that arise when the two hook up.

B. MELLOR

Relationships are complicated beasts and they come with highs and lows. After the initial flush of attraction comes the getting-to-know-you stage, when the two parties realize they share common interests. Problems flare when things get serious. Each side has its own baggage, expectations and annoying habits, but needs to figure out how to live and work closely with the other. And once money gets involved, well — that's when things can get really convoluted.

Robert Hughes, a postdoctoral researcher at the University of Washington in Seattle, had little idea of the challenges ahead when he first met leaders from the Hereditary Disease Foundation, a philanthropic organization in New York that focuses on curing the fatal neurological condition called Huntington's disease. Hughes' adviser asked him to attend a workshop the foundation had organized and, after spending two days talking with other scientists, Hughes was smitten. He now receives funding from the High Q Foundation in New York — a successor to the Hereditary Disease Foundation — and he finds them a gratifying but demanding partner. High Q expects its researchers to make clear contributions to the search for cures, and leaves little room for academic digressions. "But the benefit of working with High Q is that you get a sense that you're working as part of a community towards a common goal," he says. "There's a kind of excitement to that."

Biomedical researchers have long flirted with private foundations and wooed them into bankrolling their research. Many of these foundations are small entities focused on rare conditions and supported by patient advocates and donations. And in the past, such groups were happy to dole out money, giving academic scientists a relatively free reign to investigate the fundamental roots of disease.

But today the relationships between funders and scientists have become more passionate, needy and complex. Today's disease-focused foundations are becoming frustrated that basic research is not being converted into drugs for sick patients. They prefer to back 'translational' research, which aims to speed progress towards that goal and, to do this, they select research projects carefully and demand that scientists deliver concrete results on time. "The day of simply giving small grants is over. Now, you raise the money, you become educated about the science, and you take a seat at the table and become part of the conversation. It's absolutely new," says Susan Fitzpatrick, vice-president of the James S. McDonnell Foundation, based in St Louis, Missouri.

Researchers, for their part, are keen to help. But they are also attracted by government grants and the greater financial security and intellectual freedom these tend to offer. Through interviews with scientists and the numerous foundations that fund them, *Nature* gained a broad picture of the trade-offs involved in these relationships — a picture best illustrated by putting one foundation under the microscope.

Fatal attraction

Jenny Morton first became seriously interested in Huntington's disease in 1991, when she was setting up her first independent laboratory at the University of Cambridge, UK. During her postdoctoral fellowship, she had done a little research on the devastating illness, in which patients' physical and mental faculties inexorably waste away. Then, at a meeting, Morton met a woman whose husband and father-in-law had died of Huntington's and whose two children also inherited the disease. "I thought it was not only an academically interesting subject, but a terrible disease, so I decided to work on it," Morton recalls.

Morton applied for money from the Medical Research Council — the primary public funder of basic biomedical research in the United Kingdom — but was turned down. Like other rare conditions, Huntington's disease simply

doesn't have the same public priority as illnesses that claim many more victims, such as cancer. And because the potential market for treatments is so small, it is unattractive to drug and biotechnology companies.

These factors are frustrating for patients who have these rare conditions, and for the researchers who want to study them. So, in 2002, a group of private donors set up the High Q Foundation. The donors had one goal: to convert the fundamental knowledge about Huntington's disease gained over the past half-century into a cure.

Sniffing out success

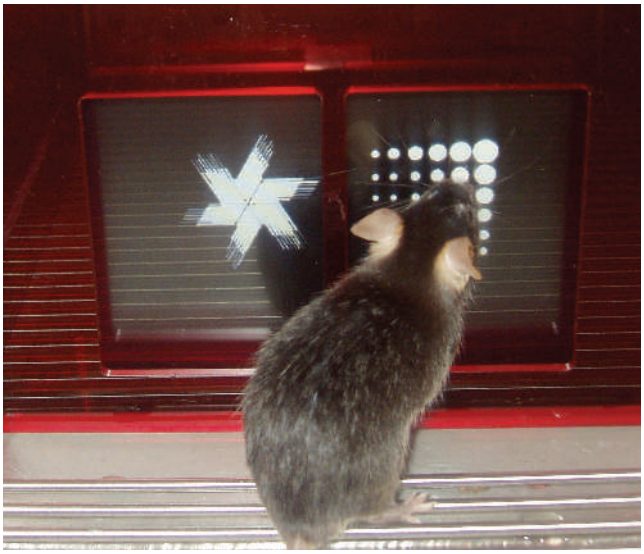
To achieve this goal, High Q and its sister group the CHDI — the successor to the Cure Huntington's Disease Initiative — now function more or less as a biotechnology company. High Q looks for disease targets — biological molecules that behave aberrantly in Huntington's. The CHDI looks for drugs that can correct the behaviour of those targets. For much of this work, the groups fund efforts in the biotechnology industry. But High Q also relies on hand-picked projects by academic researchers such as Morton to aid its quest for targets. "Ninety-five per cent of science works on the principle that the best thing to do is to let good scientists follow their noses," says Allan Tobin, a senior scientific adviser to High Q. "But this is a different attitude. We think we can direct the science."

Morton received her first funding from the Hereditary Disease Foundation in 2002, and now receives around half of her costs from High Q. The money supports studies on mice with characteristic features of Hungtington's disease. Morton also collaborates with doctors, who help her to design her animal experiments so that they are relevant to finding a cure for patients.

Morton sees many benefits to drawing much of her funding from a private foundation, rather than from a public source. For one, the organization is more willing than the government to take risks on new ideas. Five years ago, for instance, she and two other scientists brainstormed an idea for a computerized 'touch-screen' for mice. They hoped it would solve a major problem in the field by allowing them to evaluate the brain functions of animals who were physically incapable of running through a maze but could still touch a screen with their noses.

Some colleagues sneered at Morton's idea because they thought it would be impossible to teach mice with movement disorders to use such a device. Thinking that their derision would be reflected in grant reviews from government agencies, she asked High Q for the money instead, and the foundation gave her a $350,000 contract for three years to make the touch-screen prototype. Last year, Morton and her colleagues showed that they could teach mice to touch the screen and earn food pellets (A. J. Morton, T. J. Bussey & L. M. Saksida *Nature Methods* **3,** 767; 2006).

Although foundations provide valuable support for new ideas, working with these organizations can also be challenging — and most scientists still prefer to work with at least some government grants. For Morton, an important concern is the maintenance of her mice, which costs about £80,000 (US$160,000) a year. Breeding for experiments has to be planned months or even years in advance, so Morton needs to know that money will be available for those future studies and to keep staff on contracts to do them. This can be done with a typical grant from the Medical Research Council, which runs from three to five years, but High Q's awards are typically only for one year. "Short-term funding when you are trying to keep a long-term project going is very difficult," Morton says.



A. J. MORTON ET AL.

The High Q foundation supported development of a touch-screen for disabled mice.

Straight and narrow

With their new focus on delivering cures, today's foundations also insist that researchers deliver results or, in the case of high-risk projects, at least show that they tried. Some ask for detailed reports every six months or a year, and make further funding contingent on meeting strict milestones. High Q awards contracts rather than grants to ensure that scientists actually do the work they say they will do, rather than pursuing serendipitous tangents — a luxury that is allowed or even encouraged with government grants. The foundation also insists that researchers share the results of their work as quickly as possible, and discuss unpublished findings openly with colleagues at meetings.

Some researchers find this type of oversight too onerous or controlling. They resent the loss of intellectual freedom and tend to stay away from groups such as High Q. Robi Blumenstein, a former businessman who now manages the operations of High Q and the CHDI, acknowledges that their business-like procedures can chafe some academics, but he makes no apologies. "We want people to have great ideas, but we need to get them done," he says. "When we switched to this more rigorous model we acknowledged that we weren't going to get everybody to work with us."

R. PACIFICI.

"We think we can direct the science."
— Allan Tobin

Because of the constraints, researchers who do work closely with foundations like these are usually driven by something more than pure intellectual curiosity. Many scientists relish the opportunity to participate in directed, translational research that could benefit patients. Foundations may also allow them to belong to a larger research community of scientists dedicated to the same pressing problem.

In 2003, funding from High Q created the European Huntington's Disease Network, which was the first organization to unite Huntington's disease researchers across Europe and the thousands of patients that they study. "In Europe, there is no network of collaborative research like this in any other neurodegenerative disease," says Sarah Tabrizi, a doctor and researcher at University College London. This network is helping Tabrizi to find biomarkers that may predict the onset of the disease and launch clinical trials to test them. These benefits, she says, far outweigh the inconvenience of complying with short grant cycles and stringent milestones.

But for some, the expectations and sacrifices demanded by foundations are too high. For these researchers, the traditional security of government money offers a more productive union.

Tobin acknowledges that researchers need to make difficult calculations before plunging into a relationship with High Q. "We realize this is an experiment, and it involves a bunch of trade-offs," he says. "The key is that we respect each other." And that, of course, is good advice for any solid partnership — in science, as in other parts of life. ■

Erika Check is a senior reporter for *Nature* in San Francisco.

Q&A

The giving machine

Flush with Microsoft's fortune, the Bill & Melinda Gates Foundation is the largest charitable foundation in the United States. Tadataka Yamada, executive director of its Global Health Program, tells Lucy Odling-Smee how the organization aims to save lives with its wealth.

Empowering and enriching the developing world requires tens of billions of dollars a year. How can the Gates foundation hope to make a difference?

The Gates foundation contributes roughly 10% of the US$12.7 billion a year spent on health-related aid to developing countries by donors such as the United States, United Kingdom and France.

The world needs to commit a lot more funding to improving global health. Estimates of the additional resources that are needed to meet the United Nations Millennium Development Goals for health range from $25 billion to $70 billion per year. We believe that our funds have been catalytic in many ways. For example, our initial $750 million contribution to the GAVI alliance — formerly known as the Global Alliance for Vaccines and Immunisation — prompted further donations of more than $2 billion. GAVI has helped to increase immunization rates in millions of children in 70 developing countries.

How do you decide where to allocate your money within the Grand Challenges in Global Health initiative?

The Grand Challenges initiative was created to support ground-breaking research on some of the most fundamental scientific problems in global health. Its scientific board, which includes 20 scientists and public-health experts from around the world, identified 14 major challenges in global health and then reviewed proposals for research to overcome these challenges. Guided by the board's recommendations, we selected 43 projects to fund — ranging from heat-stable vaccines that don't require refrigeration to insect repellants that interfere with disease-carrying mosquitoes' sense of smell.

BILL & MELINDA GATES FOUNDATION

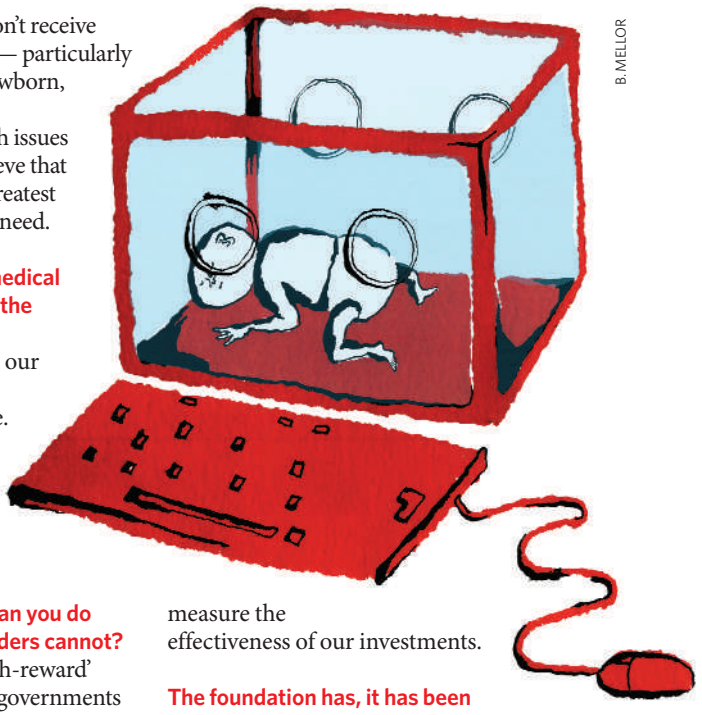
"Productivity is a tricky concept — we know that some of the research we support will fail."

Are there any areas you have specifically decided not to fund?

We focus most of our resources on the health problems that disproportionately affect developing countries and that don't receive enough attention and resources — particularly infectious diseases; maternal, newborn, and child health; and nutrition. Although many important health issues fall outside of this scope, we believe that this focus will help us have the greatest impact on the people in greatest need.

What is the best way for a biomedical researcher to win funding from the Gates Foundation?

We fund projects that fit within our global-health strategy, which is covered in detail on our website. We are very focused on real-world outcomes, so we tend to favour research that, if successful, can be translated quickly to the field.



B. MELLOR

As a private foundation, what can you do with your money that other funders cannot?

We often support 'high-risk, high-reward' projects that can be difficult for governments or the private sector to fund. For example, the Medicines for Malaria Venture, one of our major grantees, is partnering with industry to conduct clinical trials of new treatments for malaria. Private companies would not be able to fund these trials alone, because of the high financial risk of conducting expensive trials on products that do not have a market in wealthy countries.

How are scientists who are funded by the Gates Foundation held to account?

We require grantees to report on their progress against agreed-on milestones, and we often support third-party evaluations of our grants. 'Productivity' is a tricky concept — we know that some of the research we support will fail, but that doesn't mean those researchers weren't productive. Even a failed study can contribute to the body of knowledge in its field and help point the way forward. We also provide funding to outside organizations to measure the effectiveness of our investments.

The foundation has, it has been reported, no guidelines for judging the ethics of companies its funds are invested in. Is that a tenable position for the world's leading philanthropic fund?

The foundation's outside investment managers have no involvement in our grant-making decisions, and our programme teams have no involvement in investment decisions. A detailed explanation of the investment strategy for the endowment is available on our website.

Is there a risk that your dominant role in funding global health will skew the research agenda?

It is not our goal to set the global health agenda, and we can succeed only if we help encourage greater funding from other donors. It is important to note that although our resources are significant, we account for a small portion of what is spent on health globally. For example, in 2005, our global-health budget was about $1 billion, whereas the National Institutes of Health's budget was nearly $30 billion. ■

Lucy Odling-Smee is *Nature*'s associate editor for Essays.

NATURE|Vol 447|17 May 2007

BLAST IN THE PAST?

A controversial new idea suggests that a big space rock exploded on or above North America at the end of the last ice age. **Rex Dalton** reports.

The black layer in these sediments at Murray Springs may be remnants of fires after a meteorite impact.

A. WEST

Around 13,000 years ago, North America was a busy place. Millennia of ice sheets had melted away, and humans crossed from Siberia to Alaska, spreading from the Canadian woods to the lush Carolina coastline. But after just two centuries of hunting mammoth, bison and horse, this 'Clovis' culture suddenly disappeared[1], posing one of the great anthropological questions of the peopling of the Americas: why did the New World's most sophisticated hunters of the time suddenly vanish?

Now, a team of researchers is invoking an out-of-this-world cause. On 24 May, at the American Geophysical Union's meeting in Acapulco, Mexico, some two dozen scientists will present multiple studies arguing that a comet or asteroid exploded above or on the northern ice cap almost 13,000 years ago — showering debris across the continent and causing temperatures to plunge for the next millennium.

The history of geology is peppered with such notions — from the 'cosmic serpent' theory that the outer planets nudge comets onto a collision path with Earth, to the idea that an impact could have caused the collapse of Bronze-Age civilizations. Most of these theories have never become widely accepted by the scientific community. But the new team argues that its idea explains multiple observations: not only the climate cooling and the disappearance of the Clovis hunters, but also the near-simultaneous extinction of the continent's large mammals.

"The magnitude of this discovery is so important," says team member James Kennett, a palaeoceanographer at the University of California, Santa Barbara (UCSB). "It explains three of the highest-debated controversies of recent decades."

Not all will be convinced. Several leading hypotheses already explain each of these three events. A change in ocean circulation is generally thought to have brought about the onset of the millennium-long cooling, which is known as the Younger Dryas. This cooling might, in turn, have caused the Clovis hunters to disappear. And, if they had not previously been killed by disease or hunted to extinction, the big prehistoric beasts may also have been doomed by this change in climate.

The new evidence comes in the form of geochemical analysis of sedimentary layers at 25 archaeological sites across North America — 9 of them Clovis. Certain features of the layers, say the team, suggest that they contain debris formed by an extraterrestrial impact. These include spherules of glass and carbon, and amounts of the element iridium said to be too high to have originated on Earth. In addition, the rocks contain black layers of carbonized material, which the team says are the remains of wildfires that swept across the continent after the impact.

Caught knapping: were early Americans killed by a space rock?

P. DEGGINGER/ALAMY

Other experts are waiting to see how the data pan out. Vance Haynes, an archaeologist at the University of Arizona who has studied the Clovis people for more than 40 years, says the new theory could be viable. "If their geological analysis can be replicated by another group," he says, "it would make it believable."

"Their impact theory shouldn't be dismissed; it deserves further investigation," says Jeff Severinghaus, a palaeoclimatologist at the Scripps Institution of Oceanography in La Jolla, California, who studies ice cores. But he thinks it "very, very unlikely that such an event could have led to climate change", scepticism based on analysis of ice cores from Greenland, which show that some cooling had begun earlier[2].

A rocky road

Duncan Steel, an Australian asteroid expert, has seen many groups try to connect impacts with major cultural changes. In this case, he says, "the researchers make at least a prima facie case for a link."

Team members know they have a lot of scepticism to overcome. Many of them have signed on tentatively, saying they find the idea intriguing if not yet compelling. When Kennett heard of the theory, he says he thought: "This really makes sense. But it needs to be examined carefully."

Earlier theories about what caused the Younger Dryas held that it was a temporary

reversal of the global warming trend that brought the Clovis people to the Americas in the first place, across a land bridge from Siberia. At that time, North America's climate was warming and its land was being revealed by the melting of a massive ice cap that at its extreme, about 22,000 years ago, reached into what is now the heartland of the United States. Meltwater running off the ice cap fed a huge body of freshwater known as Lake Agassiz, which stretched across the upper Midwest.

At some point, the idea goes, Lake Agassiz breached a natural dam and began rushing into the North Atlantic. The rapid addition of freshwater effectively slowed down the ocean conveyor system that transports heat north from the mid-latitudes[3]. The climate cooled for about a millennium, then gradually warmed again as the Atlantic returned to more normal conditions.

Fire and ice

In place of the Lake Agassiz breach, the new theory estimates that an object up to 5 kilometres across hit the northern ice cap, causing melting and the flood into the North Atlantic. No obvious crater remains behind — perhaps, suggest those proposing the theory, because the space rock exploded in the air, or because the ice cap was thick enough to mask the impact.

The theory came together over several years, evolving with a number of twists. In Michigan in the late 1990s, archaeologist William Topping was working on a Clovis campsite known as Gainey. Topping had been exploring a theory that the Younger Dryas was caused by a nearby supernova, or exploding star. In 2001, he and Richard Firestone, a nuclear physicist at Lawrence Berkeley National Laboratory in California, published an article to that effect with data from the sediments at Gainey[4].

Critics immediately questioned the work, and after further analysis Firestone withdrew the claim. But Topping, still wedded to the supernova idea, disagreed; the two parted ways thereafter.

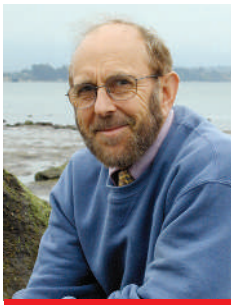
Firestone eventually joined forces with Kennett, but both give primary credit to their colleague Allen West, a self-taught geophysicist in Arizona who collected many of the new samples. "West is the champion," says Kennett.

Today, Kennett groans at the mention of the since-corrected 2001 article, but says he is fully committed to the quality of the new studies. "I'm not going to jump on some project that's not supported by sufficient data," he says. "And this project is data-driven."

Some of the data come from a Clovis site known as Murray Springs, located in southeastern Arizona and long-studied by Haynes. The sediments there include the 'black mat' of carbon layers laid down just before the onset of the Younger Dryas. Beneath that layer, West found samples of carbon spherules, ranging from 0.15–2.5 millimetres across, some of which were hollow. Firestone and West argue that similar spherules have been found at a crater in Germany, and could be remnants of an impact.

The team also retrieved samples of glassy-looking carbon, with textures they say suggest melting during an impact, as well as layers enriched in iridium, an element not found in abundance on Earth. And team member Luann Becker, of UCSB, who was previously involved in controversial claims that an asteroid caused the Permian-Triassic extinction about 250 million years ago[5], says in an abstract for the Acapulco meeting that she has found fullerenes in layers at Clovis sites — possible support for an impact event.

Team members say that they have found at least some of these markers in 25 widely scattered sites across the continent. Kennett, for instance, found spherules in 12,900-year-old sediments on Santa Rosa Island, which lies off the Californian coast and is the site of the oldest human bones found so far in North America. Other potential evidence for an impact at the time comes from neighbouring San Miguel Island. "These gave us the best examples from the West Coast," says Kennett.



"This discovery explains three of the highest-debated controversies of recent decades." — James Kennett

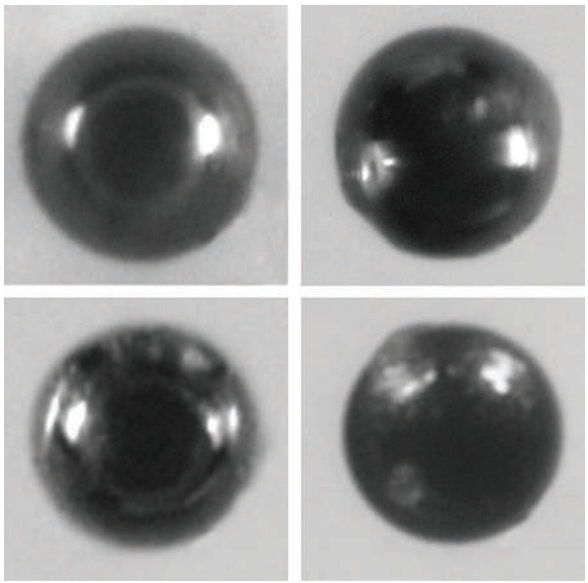
In the east, the team scrutinized a series of circular depressions that run from Georgia to Delaware. Known as the Carolina Bays, the way in which they were created has been the subject of much debate; impacts are one idea among many. Firestone thinks the bays may be craters left by debris from a Clovis-era explosion, although others remain to be convinced.

Decline and fall

Near the bays, in South Carolina, the team has also been looking for evidence at a Clovis site known as Topper. Clovis points are abundant in this region. But Albert Goodyear, an archaeologist at the University of South Carolina at Columbia, will report in Acapulco that there was a sharp decline in Clovis points at Topper during the Younger Dryas. This, the team argues, is evidence that at that time humans went through some sort of population collapse. But other archaeologists say they have no evidence of a similar decline in other Palaeoindian populations; even as the Clovis culture was disappearing, other cultures arose in its place, for reasons not entirely understood

In the end, the Acapulco meeting may cause other scientists to re-examine their evidence for what was happening in North America at the end of the last ice age. For instance, Paul Mayewski, an ice-core expert at the University of Maine in Orono, is investigating ice cores from Greenland that show a massive burning event occurred around the start of the Younger Dryas. Others will undoubtedly start digging through whatever records they have of the time. And, at the end of the Acapulco session, the team will host a dinner — throwing open the doors for critics and supporters alike to begin talking through the theory. It will be interesting to see where it goes. ■

Rex Dalton is *Nature*'s US West Coast correspondent.



A. WEST

Magnetic microspherules, reportedly evidence of an extraterrestrial impact, have been found at sites across North America.

1. Waters, M. R. & Stafford, T. W. *Science* **315,** 1122–1126 (2007).
2. Stuiver, M., Grootes, P. M. & Braziunas, T. F. *Quatern. Res.* **44,** 341–354 (1995).
3. Broecker, W. S. *Science* **312,** 1146–1148 (2006).
4. Firestone, R. B. & Topping, W. *Mammoth Trumpet* **16**(2), 9–16 (2001).
5. Dalton, R. *Nature* **431,** 1027 (2004).

Expand free journal project so poor countries can share their valuable climate data

SIR — I warmly approve your Editorial 'Millennium development holes' (*Nature* **446,** 347; 2007) about the lack of weather data from Africa and other developing countries. A further problem is that when measurements have been taken they are often not disseminated to interested organizations within their own country, let alone beyond it.

Both aspects became very apparent at the second international conference on coastal zones in sub-Saharan Africa held in Ghana in 2005 (see www.acops.org/CoZSSA/CoZSSA_conf_report_Jan06.pdf). Excellent data taken by Ghana's meteorological service along the coast, showing steadily rising temperatures and decling rainfall over 20 years, are not widely known even at the African Centre of Meteorological Application for Development at Niamey in Niger. I found a similar situation in the West Indies. These local time series show the seriousness of the problem of climate change for these countries.

There is currently no financial or other incentive to share these data. African colleagues complain that, even if they send the data to international centres, they cannot benefit, as they do not receive current issues of the journals and bulletins where the results are published.

One way forward, which I have been pursuing by lobbying UK ministers and others, is to ensure that the latest publications of such literature are sent, at no cost, to the regional and national meteorological services that are providing data in developing countries. The UN Food and Agricultural Organisation is already providing current literature to some agricultural centres in the world's poorest countries, through its AGORA programme (www.aginternetwork.org/en). The OARE programme (www.oaresciences.org/en), launched last November, has similar arrangements for the environmental-science literature, including weather and climate journals — and more countries are being included in the programme next year.

These are suitable projects for extension to more countries, and for further donations from environmental and other charities. The media organizations that focus on ghoulish pictures of climatic devastation around the world might also contribute.

Julian Hunt
Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, and House of Lords, London SW1A 0PW, UK

Primary research papers and other content in *Nature* and all Nature Publishing Group journals are made freely available online to readers in countries that are members of AGORA, OARE or HINARI (www.who.int/hinari/en), which covers health. These provide information in a timely fashion to people who might not otherwise be able to obtain it or obtain it promptly — Editor, *Nature*.

Animal-welfare section in papers would be a burden

SIR — Victoria Buck in Correspondence ('Who will start the 3Rs ball rolling for animal welfare?' *Nature* **446,** 856; 2007) calls for journals to include an animal-welfare category in the methods section of papers describing research on live animals. I disagree.

We scientists have far too many things to do to add yet another bureaucratic burden to writing papers for no useful reason. I agree that sharing information about the way animals are treated and handled during experiments could be useful, but that can and should be done in another forum.

We pay expensive rates for our animal-care facilities and personnel, and are quite often stymied by the countless new rules and regulations, many of which serve no real useful purpose other than making us jump through more hoops. We are almost regulated to inaction.

It is time for scientists to stand up and say enough is enough, even if it bucks the trend, so we can get on with our work.

C. Jimenez
Department of Physiology and Pharmacology, Oregon Health and Science University, Portland 97239, Oregon, USA

Recognition could support a science code of conduct

SIR — Recent instances of scientifically unethical behaviour such as that of Woo Suk Hwang (see *Nature* **439,** 122–123; 2006) have put pressure on governments to take official measures. In Japan, for example, a data-falsification scandal shook the scientific community last year (see *Nature* **439,** 514; 2006). In response, the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT), together with the Science Council of Japan, has decided to implement a code of conduct for scientists to detect and punish unethical acts: see www.scj.go.jp/ja/info/kohyo/pdf/kohyo-20-s3e-1.pdf.

Like the Hippocratic oath for physicians, the application of such a code to all scientific disciplines would surely be beneficial. It would make young researchers aware of the necessity of adopting ethical behaviour in the conduct of their work and would provide guidance on how to do so. Yet such misconduct must often stem from the ubiquitous pressure exerted on scientists to publish quickly and, if possible, in high-impact journals in order to have a career. The possibility of publishing a ground-breaking study depends on the quality and originality of the data. It can, therefore, become tempting to modify a few things here and there in a data set.

In this regard, adoption of a scientific code of conduct may not be enough. Efforts must be made in parallel to counteract the 'publish or perish' dogma. If there were a method for recognizing the value of a piece of work through the examination of its contribution to knowledge, rather than through the prestige of the journal in which it was published, this would be a good start.

Yan Ropert-Coudert
National Institute of Polar Research, 1-9-10 Kaga, Itabashi-ku, Tokyo 173-8515, Japan

Ground-breaking stem-cell work has been reproduced

SIR — Your News Feature 'The hard copy'[1] describes the difficulties some researchers have encountered in reproducing several ground-breaking, high-profile publications in the stem-cell field. Although your News Feature accurately summarizes our principal findings[2], the failure of a single group to reproduce our work could lead readers to believe this work has not been reproduced.

At least three independent groups have replicated the primary tenet of our paper, that neural tissue can transdifferentiate into haematopoietic cell types. Indeed, our results have been extended by showing that both human neural stem cells[3,4] and rodent olfactory stem cells[5] retain this capability. Transplantation into secondary recipients demonstrates that human neural stem-cell transdifferentiation can occur in a large animal model with long-term engraftment, similar to the finding in mice that we initially reported in our paper.

Angelo L. Vescovi*, Brent A. Reynolds†, Rodney L. Rietze†, Christopher Bjornson‡
*Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, I-20126 Milano, Italy
†Queensland Brain Institute, University of Queensland, Brisbane, Queensland 4072, Australia
‡Department of Neurology, Stanford University, California 94305, USA

1. Check, E. *Nature* **446,** 485–486 (2007).
2. Bjornson, C. R., Rietze, R. L., Reynolds, B. A., Magli, M. C. & Vescovi, A. L. *Science* **283,** 534–537 (1999).
3. Almeida-Porada, G. *et al. Brit. J. Haematol.* **130,** 276–283 (2005).
4. Shih, C. C. *et al. Blood* **98,** 2412–2422 (2001).
5. Murrell, W. *et al. Dev. Dyn.* **233,** 496–515 (2005).

Published contributions are edited.

BOOKS & ARTS

The search for novelty

Evo-devo is not the first attempt to understand how evolutionary innovations arise.

From Embryology to Evo-Devo: A History of Developmental Evolution
edited by Manfred Laubichler & Jane Maienschein
MIT Press: 2007. 569 pp. $55, £33.95

Wallace Arthur

Those of us lucky enough to have become, by whatever route, students of the interdisciplinary field of 'evo-devo' have a sense of history in the making — a feeling that a new synthesis is being born in biology, even if the birth is painfully slow. Indeed, many would say it is a rebirth: of comparative embryology in the age of the gene. The original birth, at the hands of Karl Ernst von Baer and Ernst Haeckel in the nineteenth century, and its late-twentieth-century equivalent, with Stephen Jay Gould's book *Ontogeny and Phylogeny* and the discovery of the homeobox, are connected by much intermediate work. This is made clear in *From Embryology to Evo–Devo*, edited by Manfred Laubichler and Jane Maienschein. The book tells the history of this work, which has not been told prominently enough before.

Both the book and its title derive from a pair of linked events: a seminar held at Woods Hole, Massachusetts, in 2001, and a workshop in 2002 organized by the Dibner Institute. These gatherings were mostly attended by historians of science, but also included a few practitioners of evo-devo. In the book's final chapter, one of the latter, Günter Wagner, describes this mix memorably as "birds and ornithologists together".

My first reaction to the book's title was that evo-devo has many roots, of which embryology is just one. But as Laubichler and Maienschein explain in their introductory chapter, "embryology and evo-devo are merely historical markers for the late nineteenth and the early twenty-first century, respectively". In fact, this first chapter is essential reading, as it outlines the structure of the book, which would otherwise be somewhat inscrutable. To my mind the flow is a bit odd: to have the first section defined by a block of historical time (the early twentieth century) but the other two sections defined in different (non-chronological) ways seems rather strained. So I view the book overall as something of a haystack, but it does contain many valuable needles.

My problem here is that a book of 16 chapters, some consisting of more than 50 pages (too long!), by assorted authors, is too big to do justice to in a short review. So I will concentrate on a few things that struck me as important recurring themes.

PRIVATE COLLECTION/BRIDGEMAN ART LIBRARY

Prize Monet: finding the roots of evo-devo can be like searching for a needle in a haystack.

First, history is more complex than scientists often think, and consists of the interactions of a multitude of figures, not just those who show up repeatedly in the popular literature. I found the chapters on less-well-known figures particularly interesting, such as Alan Love's chapter on Dwight Davis and William Gregory. Although Davis was a morphologist and Gregory a palaeontologist, they were both interested in the origin of evolutionary novelties. And William Wimsatt provides an excellent comparison of the very different backgrounds of Rupert Riedl and Lancelot Law Whyte. He contrasts this with the similarities of their focal area, the 'internal' side of evolution (Riedl's concept of 'burden' and Whyte's 'internal selection').

Second, Haeckel's idea of recapitulation — the fleeting occurrence in embryos of 'advanced' animals of certain features of their ancestors — is not dead, despite many reports to the contrary. This point is made at some length by Frederick Churchill, who distinguishes between strong and weak versions of recapitulation. He notes that there are intermediates between them, and that they have changed over time, with the weak version becoming "even milder". Churchill concludes that only the strong version of recapitulation met "a natural death", and I think he is right. It all hinges on what is meant by 'recapitulation'. If it means 'repeats exactly' then of course it is false. On the other hand, if it means that the ontogenies of derived animals often retain features of the ontogeny of an extinct ancestor (as in the case of gill clefts in mammalian embryos), it is self-evidently true, but is more a statistical pattern than an inviolable law.

Third, and most important in my view, the origin of novelty is becoming one of the major themes of evo-devo. Attention is shifting from the retention of the old (as in recapitulation) to the creation of the new (be it an eye, a leg, a feather or even a whole body plan). Both the historical and the current importance of novelty emerge repeatedly in the book.

How do novelties arise? We can't yet agree on a definition for them, let alone answer this fundamental question. But we can see the nature of the challenge ahead. Wagner points out that there is a growing connection between microevolutionary (intraspecific) evo-devo and quantitative genetics (where intraspecific

variation is analysed in terms of quantitative trait loci). This connection is a positive thing, although it is perhaps limited in scope because it may not solve what many perceive as the *raison d'être* of evo-devo. As Wagner says: "One of the main sources of intellectual excitement in devo-evo *(sic)* is the prospect of understanding major evolutionary transformations." Whether these end up being unique events or long-term accumulations of the mundane remains to be seen, but either answer will be exciting in its own way.

Overall, then, the book is a mixed bag, but contains many important contributions to the past, present and possible future of evo-devo. It is definitely a reference book rather than something to read from cover to cover. Its 'haystack' nature is off-putting at first, but the best strategy is to dive in (with the aid of the short introductory chapter) and see what you can find — including those needles that space has not permitted me to discuss. ■

Wallace Arthur is European editor of the journal *Evolution & Development*. He is in the Department of Zoology, National University of Ireland, Galway, Ireland.

A singular view of ageing

Aging of the Genome: The Dual Role of DNA in Life and Death
by Jan Vijg
Oxford University Press: 2007. 372 pp.
£70, $114.50 (hbk); £32.50, $65 (pbk)

Linda Partridge

There is no shortage of theories of ageing. Confronted by the terrifying realization of mortality, human ingenuity has created an interesting array of explanations, including toxins produced by gut bacteria (curable by eating yoghurt) and reduced secretions from the testicles (curable by transplants of testicular tissue from monkeys). There is now general agreement that ageing is caused by the accumulation of damage. Key issues are the exact types of damage responsible for functional impairment and death, and the processes that generate this damage and protect against it. Jan Vijg's excellent book *Aging of the Genome* makes no concession of equal space for the many candidates subject to current scrutiny. Rather, it critically examines the case for one — somatic mutation.

First formulated in the 1950s, this theory suggests a key role in ageing for the accumulation of random alterations to DNA in somatic tissues (all tissues other than the reproductive germline cells). DNA is being constantly bombarded with chemical and physical challenges that induce random alterations, including structural damage and changes in nucleotide sequence and organization. But unlike other biomolecules, such as proteins and lipids, the damaged DNA cannot be simply broken down completely and remade, because it holds unique information. Instead, cellular pathways detect alterations and, contingent on the type of cell and the nature of the changes, this variously leads to DNA repair, arrest of the cell cycle (preventing cell division), cellular senescence or death, or toleration of the change. In some cell types, some forms of DNA alterations accumulate with age, with evidence for genomic hotspots and considerable variation between individuals. Cancer is a clear case where DNA alterations can give rise to age-related pathology; their role in other aspects of functional decline is less clear, with the exception of mutations in DNA within mitochondria, the organelles that power cells. As well as leading to ageing directly, DNA alterations could lead to ageing as a result of cellular defence mechanisms, such as selective cell death, although there is little evidence for this.

Vijg gives a clear and thoughtful account of this complex, and potentially confusing, body of work and its limitations: little work has been done on non-dividing cells; most evidence has come from cells in dishes rather than in tissues; measuring DNA alterations is difficult; a net change in levels of DNA alterations can be attributable to several different events including

EXHIBITION

A painful pleasure

Three bodies writhe in agony. Their limbs are distorted, their features unrecognizable, their entrails burst out. The *Crucifixion* triptych by Francis Bacon (the central panel of which is shown here) had no religious meaning for the painter, for whom the work was simply about the expression of extreme sensation. In Tiepolo's painting of the martyrdom of Saint Agatha, the young woman's ecstatic gaze is thrown heavenwards as she awaits the blow of her tormenter's sword. These two paintings are the artistic highlights of the exhibition *Schmerz* (*Pain*), which runs until 5 August at the Medical History Museum of Humboldt University and the neighbouring Hamburger Bahnhof Museum in Berlin.

The exhibition brings together artistic and medical views of pain. Opposite Bacon's *Crucifixion* is a glass cabinet containing pathological preparations of organs. Under the title 'The pleasure of pain', Tiepolo's Agatha is displayed along with forensic photographs showing fatal accidents that occurred during masochistic sex. The borders between art and documentation begin to blur, which makes the exhibits all the more disturbing. Video interviews with people who self-mutilate, by German film director Valenska Griesebach, could easily be from the files of a psychiatrist. And what differentiates a display in a vitrine from a pathology lab or an art installation?

The exhibition aims to show pain in all its forms, rather than to understand it, and plays with superficial similarities between different depictions. A video by Bruce Naumanns in which a violin string is repeatedly plucked appears next to chattering patch-clamp recordings in the only exhibit that gives a nod to neurophysiological research on pain. That's not enough to justify the exhibition's claim to build a bridge between science and art. Rather, wandering through this labyrinth of abominations, the question that comes most immediately to mind is why Christianity really needed to glorify this most ugly of all human sensations into the pinnacle of mystical experience. Stefan Klein

PINAKOTHEK DER MODERNE, MUNICH

cell death; and pinning functional decline to altered DNA is a major challenge.

How can we tell if a theory of ageing is correct? If one underlying process is key to loss of function and death, then slowing it down should also slow down ageing. Vijg points to excitement about one successful application of this approach. The trail-blazing work of Michael Klass, who isolated the first long-lived mutant animals in the nematode worm *Caenorhabditis elegans*, culminated in the discovery of evolutionarily conserved mechanisms for increasing healthy life span. A critical test of a key role for DNA alteration in ageing, would therefore be to reduce it, by increasing the activity of pathways that combat it. For biochemical reasons, this is going to be difficult, however. Vijg gives a fascinating account of single-gene mutations, such as those causing Werner's syndrome, that seem to accelerate some aspects of ageing in humans. Progerias such as Werner's syndrome involve lesions in pathways that sense and repair alterations in DNA, suggesting that these processes are crucial for assuring a normal life span. More telling evidence for a key role for DNA alterations in normal ageing comes from the finding that some types of change in DNA accumulate more slowly in mutant long-lived mice, or when life span is extended by dietary restriction.

If this clear and insightful text has a cloudy patch, it is the account of the evolutionary basis of ageing. Contrary to Vijg's account, a solid mathematical foundation in theoretical population genetics and dynamics has revealed that, despite the loss of fitness it causes, ageing can evolve, through just two routes. Germline mutations with deleterious effects at later ages in adulthood, such as Huntington's disease in humans, can accumulate in populations through mutation pressure. Mutations that cause a benefit earlier in life, such as high fecundity, but at the cost of a higher subsequent rate of ageing, can also enter populations, by natural selection. Vijg points out that detection and repair of DNA alterations is costly, potentially leading to a trade-off between reproduction and somatic maintenance. Mutation pressure could also be important; small, life-long effects on maintenance of the integrity of DNA could lead to variation in the rate of ageing.

This is a work of real scholarship, a critical account of a huge swathe of work with no fewer than 879 references. It will nonetheless be enjoyable for non-specialists, and the opening chapters are a brisk walk through much of modern biology. The sharp focus on one type of damage, the excellent writing style and the well argued, personal perspective of the author contrive to keep the reader going. As Vijg acknowledges, the jury is out on the role of alterations to DNA in ageing, and this text points the way to the kind of research that will be needed to resolve the issue. ■

Linda Partridge is in the Department of Biology, University College London, Gower Street, London WC1E 6BT, UK.

A man with prosopagnosia uses a grid to appreciate his girlfriend's face in the film *In Vivid Detail*.

D. BRATT

FILM

An unfamiliar face

In Vivid Detail
directed by Dara Bratt
presented by the Alfred P. Sloan Foundation at the Tribeca Film Festival, New York, in spring 2007.

Emma Marris

Prosopagnosia is a neurological condition often referred to as 'face blindness'. Individuals with this condition cannot recognize other people by their faces, which look about as uniform to them as two stones do to everyone else. It is a disorder with rich cinematic implications, as film stars often have beautiful or glamorous faces, and films rely on the audience's ability to identify characters by their faces.

In the short film *In Vivid Detail*, a man falls in love with a conventionally pretty woman, and she reciprocates, only to feel snubbed when he apparently ignores her when she has put her hair up and changed her shoes. He explains his condition, and she must decide whether she can date a man who can't really 'see' her face. In the end, he makes a special effort to show that he can comprehend her face, albeit in a different way to most people.

The man's disorder is revealed in subtle ways before he explains it. Talking about two chocolate-shop owners, he tells his new love "Those guys are brothers," when it is obvious to the viewer that they are identical twins. And the film almost overdoes it with scenes that reveal his excellent eyesight and attention to detail in other areas.

Student film-maker Dara Bratt spoke with scientists and prosopagnosia patients to do her research, but she admits that she is not sure whether the final conceit, in which the man comes to understand the woman's face by sketching it in a grid and appreciating it as an abstract pattern, would work for people with the condition.

Bratt says she chose a conventional beauty to play the woman so the character would be more startled and upset when the man fails to respond to her pretty face. But this casting choice also puts the audience in the same precarious position of potentially failing to recognize her when she changes her accessories and hairstyle — she could be just another generically pretty blonde. The man's inability to 'see' faces might have been more poignant if the woman's face had more character.

Short films are often used as calling cards in the industry, screened at festivals to introduce a film-maker's work. *In Vivid Detail* was premiered in this year's Tribeca Film Festival in New York. Bratt says she hopes to do more work on neurological topics, and is already working on feature-length ideas. ■

http://www.tribecafilmfestival.org

Emma Marris is a reporter for *Nature* based in Washington DC.

NEWS & VIEWS

REGENERATIVE BIOLOGY

New hair from healing wounds

Cheng-Ming Chuong

In mammals, most wounds heal by repair, not regeneration. It now seems that, as they heal, open skin wounds in adult mice form new hair follicles that follow similar developmental paths to those of embryos.

A person whose leg is amputated is left with a stump, whereas some amphibians have the awe-inspiring ability to regenerate new limbs after amputation. Overall, adult mammals have very limited regenerative ability; this could be due to a lack of stem cells or the absence of proper environmental signals. On page 316 of this issue, Ito *et al.*[1] report an unexpected finding that could change our current understanding of repair and regeneration in adult mammals. They show that the outermost skin layer — the epidermis — of wounded adult mice can regenerate new hair follicles during healing, and that this ability depends on the characteristics of the wound.

Observations some 50 years ago had indicated that, in mice, rabbits and humans[2–5], some hair follicles develop anew after wounding. But because of a lack of definitive evidence, these findings had generally been discounted. Now, using advanced cellular and molecular techniques to study normal adult mice, Ito *et al.* have rediscovered these forgotten phenomena. The authors made large wounds (1–2.25 cm^2) on the animals' backs, to the full depth of the skin. They found that if, following wound closure, the healed wound was larger than around 0.5 cm in diameter, new hairs formed at the centre of the wound. An examination of the sections of the healed skin revealed changes that resembled various stages of embryonic hair-follicle development. The new hair follicles grew, passed through the hair cycle, and eventually became indistinguishable from neighbouring hair (Fig. 1).

Why has this phenomenon previously been missed? The reason might be that large wounds in humans are treated with sutures and dressings. Although such procedures help wound closure, they might not be ideal for the generation of new hair follicles. Similarly, it is not common practice to leave wounds open in mice. The authors did this here because they wanted to trace the fate of hair-follicle stem cells, which normally reside in the bulge in the hair follicle (Fig. 1), during wound healing. Thus, a combination of altered experimental design and careful observation led to these exciting findings, which verify the preliminary observations from the 1950s and help to explain the controversy.

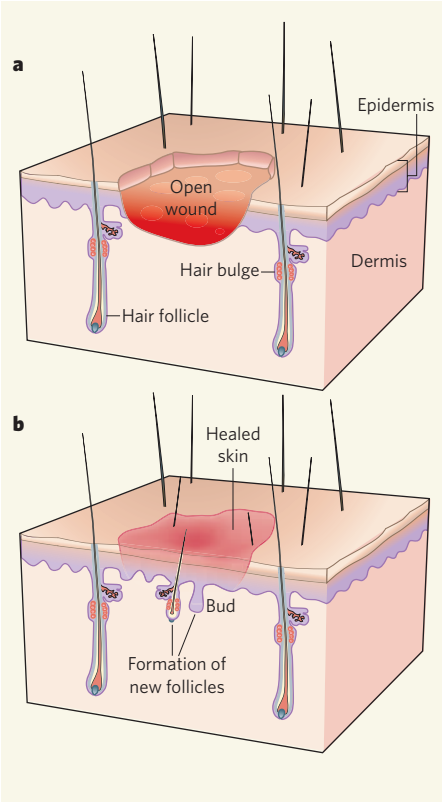


Figure 1 | Formation of new hair in a healed wound. a, Ito and colleagues[1] found that when a large open wound is generated in the skin of adult mice, re-epithelialization occurs. **b**, If the healed skin is larger than around 0.5 cm in diameter, new hair follicles originating from the epidermis form in the centre of the wound.

What is the origin of the cells that make up these new hair follicles? Are they derived from existing hair follicles located at the wound edge, or from inter-follicular epidermis? Under normal conditions, the epidermis and the hair follicles maintain separate stem-cell compartments[6–9]. Following wounding, however, cells derived from the hair bulge contribute to re-epithelialization — a process of new epidermis formation to cover the denuded dermis layer. This indicates that hair-bulge stem cells can turn into wound epidermis, although their contribution seems to be transient; with the exception of some cells from the upper part of the follicles, most of the hair-bulge-derived cells later disappear from the wounded epidermis[6–9].

To determine the origin of the hair follicles that develop following wound repair, Ito *et al.*[1] used a mouse model in which the bulge cells or inter-follicular epidermal cells were genetically labelled before wounding, so that they could be traced afterwards. The authors found that cells constituting the newly formed hair follicles are derived from inter-follicular epidermis, and not from existing hair bulges. Whether the new hair follicles themselves are generated from epidermal stem cells or through de-differentiation of existing epidermal cells is unknown.

That the epidermis can turn into skin appendages (hairs, glands, feathers) is not entirely surprising. Previous studies have shown that, by combining cells from different tissue components under appropriate experimental conditions, scales can turn into feathers, oral mucosa (the membrane covering structures inside the mouth) can turn into tooth-like appendages, and even the corneal epithelium can become hair follicles[10]. Some of these changes can be achieved by altering the balance of relevant molecules in the cell. For example, β-catenin is a component of a signalling pathway mediated by Wnt proteins that is involved in regulating development. Increasing the activity of β-catenin can trigger the formation of new hair follicles in the interfollicular epidermis of adult mice without the use of hair-bulge stem cells[11]. However, such cellular processes are occurring under experimental conditions. What is remarkable about the findings of Ito and colleagues is that skin wounds stimulate the formation of hair spontaneously, as part of the normal healing process.

Ito *et al.*[1] went on to show that wounding activates the Wnt-mediated signalling pathway, which is essential for normal hair development and cycling[12]. Inhibiting this pathway in the skin during wounding led to a substantial decrease in the number of new hairs. By contrast, when mice with increased Wnt activity in

their epidermal layer were wounded, there was a significant increase in new hair follicles compared with mice with normal Wnt activity. As these cellular events seem to recapitulate those seen in embryonic development, it is possible that hair formation during embryogenesis and following wounding share several signalling pathways, including Wnt.

What are the essential criteria for triggering the formation of new hair follicles in a patch of adult skin? The size of the healed wound seems to be critical. This implies that an 'embryonic skin-like field' must be established first. Self-organization of hair follicles then progresses, and periodically arranged primordia (aggregates of embryonic cells that indicate the first traces of a structure) emerge. As new hair patterns after wounding are not predetermined, it is possible to manipulate the number and size of the follicles through positive- and negative-feedback regulation of inhibitors and activators of signalling pathways such as Wnt[10,13].

Adult organisms contain several types of cells with remarkable regenerative potential[12,14,15] — if we could only provide the appropriate chemical and physical environment. The best teacher for this is nature. Indeed, an event that parallels the work of Ito and colleagues is the regeneration of deer antlers. After an antler is cast, the large open wound that forms is followed by re-epithelialization and the development of new hair follicles, as well as budding of the new antler[16]. Further studies on animal models should reveal other unexpected and ingenious ways of awakening stem cells with appropriate environmental cues when regeneration is needed[10].

Repair and regeneration seem to be competing processes. As closing wounds fast is essential for survival, repair often dominates. Regenerative medicine promises to identify natural healing power and a shift from repair to regeneration. Thus, by simply altering the environment of stem cells[10] during wound healing, future wounds might heal with appendages reformed. As human and mouse skin heals differently, the results of Ito *et al.*[1] are yet to be verified in humans. However, these findings will undoubtedly inspire new thinking in the management of alopecia, in tissue engineering and in the regeneration of other organs. ■

Cheng-Ming Chuong is in the Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA.
e-mail: chuong@pathfinder.usc.edu

1. Ito, M. *et al. Nature* **447,** 316–320 (2007).
2. Kligman, A. M. & Strauss, J. S. *J. Invest. Dermatol.* **27,** 19–23 (1956).
3. Billingham, R. E. & Russel, P. S. *Nature* **177,** 791–792 (1956).
4. Breedis, C. *Cancer Res.* **14,** 575–579 (1954).
5. Lacassagne, A. & Latarjet, R. *Cancer Res.* **6,** 183–188 (1946).
6. Ito, M. *et al. Nature Med.* **11,** 1351–1354 (2005).
7. Levy, V. *et al. Dev. Cell* **9,** 855–861 (2005).
8. Claudinot, S. *et al. Proc. Natl Acad. Sci. USA* **102,** 14677–14682 (2005).
9. Levy, V., Lindon, C., Zheng, Y., Harfe, B. D. & Morgan, B. A. *FASEB J.* **7,** 1358–1366 (2007).
10. Chuong, C.-M., Wu, P., Plikus, M., Jiang, T. X. & Widelitz, R. B. *Curr. Top. Dev. Biol.* **72,** 237–274 (2006).
11. Silva-Vargas, V. *et al. Dev. Cell* **9,** 121–131 (2005).
12. Fuchs, E. *Nature* **445,** 834–842 (2007).
13. Maini, P. K., Baker, R. E. & Chuong, C.-M. *Science* **314,** 1397–1398 (2006).
14. Jahoda, C. A. & Reynolds, A. J. *Lancet* **358,** 1445–1448 (2001).
15. Toma, J. G. *et al. Nature Cell Biol.* **3,** 778–784 (2001).
16. Croft, N. J. & Randall, V. A. in *Hair Science & Technology* (ed. Van Neste, D.) 69–74 (Skinterface, Tournai, Belgium, 2003).

OPTICS

Beyond diffraction

Evgenii E. Narimanov and Vladimir M. Shalaev

A material with a cunningly designed optical response overcomes a fundamental limit to image resolution. This 'hyperlens' produces magnified images of objects smaller than the wavelength of the imaging light.

When an object is illuminated, information about features smaller than the wavelength of the incident light is carried in evanescent waves, whose amplitudes decrease exponentially with distance. This rapid decay means that detail is lost when an image of the object is viewed in its far field. Two papers published in *Science*[1,2] contain experimental details of a 'hyperlens' that circumvents this problem, turning evanescent fields into propagating waves, and so producing magnified far-field images of sub-wavelength structures.

For propagating electromagnetic waves, the wavelength defines the scale over which the accompanying electromagnetic field varies. When an object is moved by a distance much smaller than this wavelength, it will be subjected to essentially the same field as before, and an image formed by the light scattered off the object will also remain unchanged. The light's wavelength therefore gives a measure of the image resolution, generally referred to as the Abbe diffraction limit[3], after the nineteenth-century German physicist Ernst Abbe.

At the heart of the hyperlens concept[4,5] lies a nanostructured 'metamaterial' whose dielectric constant — a measure of a material's response to the electric field of the incident light — has opposite signs in two orthogonal directions. The effect of this anisotropy is to do away with the lower limit on the wavelength of a propagating field that is characteristic of a conventional, isotropic medium. With no lower limit on the propagating light's wavelength, there is no diffraction limit — and so, theoretically, unbounded image resolution.

As soon as waves of very small wavelength emerge from this 'optical hyperspace' into air, however, they can no longer propagate, and again become evanescent. To deliver the sub-wavelength information carried by such waves into the far field, one must first increase their wavelength to the point when propagation in air is possible. The cylinder (or half-cylinder) geometry of the hyperlens is specifically designed to achieve this, by slowly increasing the wavelength as the field spreads away from the centre of the device (Fig. 1).

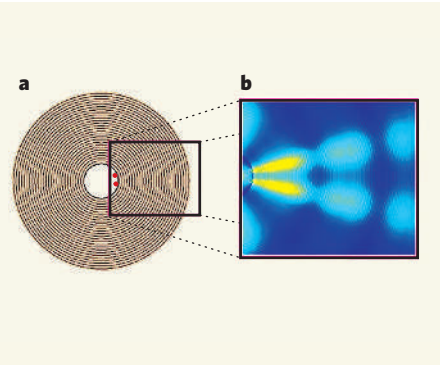


Figure 1 | How to build a hyperlens. a, As theoretically proposed[4,5], a hyperlens consists of a (half-)cylindrical layered object that has dielectric constants of different signs across the layers (radial axis) and along the layers (tangential axis). The target to be imaged (here, two point sources, red) would, in a practical realization, be illuminated from the top (into the plane of the page) or from the side, in which case half of the cylinder is cut away. **b,** Although the distance between these point sources is significantly smaller than the wavelength of the illuminating light, the effect of the anisotropic hyperlens medium is to progressively increase the distance between their 'images' until, at the outer surface, it is larger than the wavelength, and can be resolved by a conventional microscope. (Figure adapted from ref. 4.)

In their experimental realizations of the hyperlens, Liu *et al.*[1], of the University of California, Berkeley, and Smolyaninov *et al.*[2], of the University of Maryland, Baltimore, use a half-cylinder[1] and a cylinder[2] of layered metamaterials whose dielectric constant is strongly isotropic in the radial and tangential directions. The objects to be imaged are placed in the hollow middle and illuminated from the outside — in Smolyaninov and colleagues' case, from the top of the cylinder, and for Liu and colleagues from the side of the half-cylinder.

The two groups differ in their methods for achieving the necessary strong anisotropy in their hyperlens medium. Liu *et al.*[1] use a curved,

periodic stack of silver and aluminium oxide deposited on their half-cylinder cavity, which is fabricated on a quartz substrate. They use this system to produce an image of a pair of lines with sub-wavelength spacing and the letters 'ON', both with sub-diffraction resolution.

Smolyaninov *et al.*[2], on the other hand, combine the idea of a hyperlens with the earlier concept of the superlens[6]. A superlens is made of a single layer of material with a negative refractive index. Rather than immediately converting evanescent waves into propagating fields, a superlens enhances the evanescent waves through resonant coupling to fields on the surface of the lens known as surface plasmon polaritons. In Smolyaninov and colleagues' device, formed of concentric rings of the plastic polymethyl methacrylate (PMMA) on a gold surface, the evanescent waves experience such a boost, in addition to the device's strong anisotropy. The magnifying action of this lens is demonstrated by imaging rows of two or three PMMA dots placed near the inner ring of the device.

These novel imaging devices[1,2] have significant advantages over traditional approaches, and should therefore find numerous applications in optical imaging. Above all, they produce a direct optical image, and so do not require the time-consuming scanning process of near-field scanning optical microscopy.

Recently, the Berkeley group has also developed an alternative approach to the far-field superlens[7], in which the conversion from evanescent to propagating waves is achieved through scattering on surface corrugations. Although it is still too early to tell which of these devices will prove the most useful, the recent explosive progress in sub-wavelength imaging undoubtedly opens up many exciting possibilities. ■

Evgenii E. Narimanov and Vladimir M. Shalaev are in the Department of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907, USA.
e-mails: evgenii@purdue.edu; shalaev@purdue.edu

1. Liu, Z., Lee, H., Xiong, Y., Sun, C. & Zhang, X. *Science* **315,** 1686 (2007).
2. Smolyaninov, I. I., Hung, Y.-J. & Davis, C. C. *Science* **315,** 1699–1701 (2007).
3. Abbe, E. *Arch. Mikrosk. Anat. EntwMech.* **9,** 413–468 (1873).
4. Jacob, Z., Alekseyev, L. V. & Narimanov, E. *Opt. Express* **14,** 8247–8256 (2006).
5. Salandrino, A. & Engheta, N. *Phys. Rev. B* **74,** 075103 (2006).
6. Pendry, J. B. *Phys. Rev. Lett.* **85,** 3966–3969 (2000).
7. Liu, Z. *et al. Nano Lett.* **7,** 403–408 (2007).

ECONOMIC ECOLOGY

In the market for minke whales

Stephen R. Palumbi

The capture–recapture technique is a mainstay of ecology. This principle has been applied with individual genotyping to estimate how many accidentally killed minke whales reach the markets of South Korea.

Behind this stack of parboiled strips of fatty skin, and red meat with thick strips of blubber (Fig. 1), lies a story of scientific detective work. The picture shows a shop in South Korea where the whale products for sale could have come from an endangered population of minke whales from the East Sea, elsewhere called the Sea of Japan. Despite the ban on commercial whaling administered by the International Whaling Commission (IWC), this sale is probably perfectly legal: whales accidentally entangled and killed by fishing gear in the East Sea can be sold for food, as long as the accidental death is reported.

But how many whales really suffer this fate? Writing in *Molecular Ecology*, Baker and colleagues[1] describe how they have developed a genetic capture–recapture method to tackle this question. They conclude that some 827 whales were killed and sold this way during their five-year study — far more than the 458 reported to the IWC.

Minke whales in the East Sea make up a distinct population referred to by the IWC as the J stock, which suffered severe declines until the moratorium on commercial whaling began in 1986. Despite the ban, meat from this population continued to be found in Asian markets[2], and these animals were eventually traced to whales accidentally drowned by fishing gear. After the genetic detection of J-stock animals in Japanese markets, Japan and Korea began reporting whales that were mistakenly killed — up to 150 in some years, but usually far fewer. The reported value of a minke whale (at least US$30,000 each) underscores the difficulty of protecting this population from exploitation[3].

Baker *et al.*[1] took the approach of market ecologists, setting up a genetic capture–recapture study. Mark–recapture surveys in ecology seek to estimate the number of animals in a population by releasing a known number of marked individuals, and then estimating total population size based on the fraction of marked individuals recaptured[4]. Capture–recapture methods use individual identification tools such as DNA genotype analysis instead of marks[5]. And like a typical experiment in ecology, the chance of identifying the same whale twice in a market depends on the whale's 'lifespan' in the market — how long meat samples from an individual whale last before they are all sold. To estimate this figure, a co-author, Justin Cooke, developed

N. BEHRING-CHISHOLM/GREENPEACE

Figure 1 | Whale for sale. This parboiled minke-whale meat from the East Sea (Sea of Japan) is on offer at a shop in South Korea. A five-year estimate[1] based on individual genetic identification of whale products in South Korea shows that almost twice the number of minke whales were accidentally killed in fishing gear compared with official estimates. Especially given the large number of minkes from the same population that are accidentally killed and landed in Japan, this rate of loss is likely to be unsustainable.

a recapture model that considers that a whale persists on the market with a particular half-life. The model takes observations of identical DNA genotypes in the same and different surveys, and estimates the flow of whales into the market and how long each one persists.

Individual identification was based on genotypes at multiple genetic loci. The regulations laid down by the Convention on International Trade in Endangered Species require permits to transfer whale-meat samples or DNA out of the country in which the whale was landed. So initial work had to take place in hotel rooms using a portable lab. An innovation in this study was the use of multi-gene microsatellite amplification methods, which allowed the archiving of a complete, synthetic set of gene markers from each individual whale for study elsewhere.

Baker and colleagues collected 289 samples in 12 surveys in South Korean markets between 1999 and 2003. Sometimes they found the same whale individual twice in one survey, but this occurred only 13% of the time. One individual was discovered four times in one survey. These low rediscovery rates during the same survey suggest that there are many different individual whales in the market at any one time. In nine cases the authors found the same individual in two surveys — always in subsequent trips. And once they found an individual in three consecutive surveys.

Fitting these data into the capture–recapture model, Baker *et al.*[1] conclude that minke whales in the Korean market are sold quickly. The half-life of a whale — the time during which half of the whale's meat is estimated to have been sold — is about 1.8 months. The only previous hint about the market half-life of whale products came from Japan and was far longer; products from a genetically unique fin/blue whale hybrid were purchased in Japan in 1993 and 1995, but came from an animal killed off Iceland on 29 June 1989, during the last scientific hunt of fin whales[6].

Poorly known features of Korean whale markets could affect the accuracy of the authors' estimates[1]: in particular, variance in the number of products from different whales, differences in half-life owing to different storage methods, or non-random distribution of products among markets. Baker *et al.* suggest that these factors would probably make their figures lower bounds. But an overestimate of whale flow might occur if products from a particular whale individual were sold to only one or a few vendors. This possibility can be investigated only with more complete sampling of single markets. A future refinement is that truly accurate estimates will probably benefit from more frequent sampling.

Although the current model is simple, Baker *et al.* argue that the value they calculate probably represents a minimum for the number of J-stock whales in the East Sea that are accidentally killed and sold in South Korea. Adding the figure of 827 to the 390 J-stock minkes that Japan reported killed accidentally in the Sea of Japan by their fishermen in 1999–2003 gives a total (assuming no under-reporting in Japan) of more than 1,200 whales taken from this protected stock during this period.

Models of minke-whale population dynamics suggest that the J stock cannot sustain this rate of loss. An earlier model, based on the current best models used by the IWC, concluded that killing 50–150 animals a year would drive the population to near extinction by mid-century[2]. Continued take of more than 200 animals a year will accelerate this collapse. Populations of minke whales in the offshore waters of the western North Pacific are larger than the J stock, and these whales are hunted under a permit that the Japanese government issues to itself every year to conduct lethal scientific research. The larger oceanic population might be able to sustain a catch of 200 animals a year, but the structure of whale populations is sometimes so local that small and isolated populations such as the J stock cannot support a loss rate that may seem minor on a whole-ocean scale.

Debate about a return to commercial whaling often runs up against the lack of clear information about the number of whales actually hunted, and the size of the local population from which those animals are drawn. The work by Baker *et al.*[1] takes forensic identification of whale products[7] to the individual level and provides a market-eye view of commerce in whale meat that would be unavailable from official reports. For J-stock minke whales, it shows that reform of rules for the legal sale of accidentally killed animals is required if the East Sea is to be the home of whales in the future. ■

Stephen R. Palumbi is in the Department of Biological Sciences, Stanford University, Hopkins Marine Station, Pacific Grove, California 93950, USA.

1. Baker, C. S. *et al. Mol. Ecol.* doi: 10.1111/j.1365-294X.2007.03317.x (2007).
2. Baker, C. S. *et al. Proc. R. Soc. Lond. B* **267,** 1191–1199 (2000).
3. Clapham, P. & Van Waerebeek, K. *Mol. Ecol.* (in the press).
4. Williams, B., Nichols, J. & Conroy, M. *Analysis and Management of Animal Populations* (Academic, New York, 2002).
5 Mills, L. S. *et al. Ecol. Appl.* **10,** 283–294 (2000).
6. Cipriano, F. & Palumbi, S. R. *Nature* **397,** 307–308 (1999).
7. Baker, C. S., Cipriano, F. & Palumbi, S. R. *Mol. Ecol.* **5,** 671–685 (1996).

SPINTRONICS

Silicon twists

Igor Žutić and Jaroslav Fabian

For decades, silicon has been the dominant material for conventional, charge-based electronics. A new twist makes silicon ripe to enter the domain of spintronics, where the new currency is electron spin.

Modern computers present serious challenges for conventional, silicon-based electronics. Ever-increasing demands on processor speed, memory storage and power consumption — the era of the laptop that can keep us warm in winter is fast upon us — are forcing researchers to explore unfamiliar territory in the quest for increased performance. In these endeavours, Appelbaum and colleagues (page 295 of this issue)[1] report a possibly decisive development: the first demonstration of the transport and coherent manipulation of electron spin in silicon.

In spin electronics, or spintronics[2], information is represented by spin and by its proxy, the direction of magnetization. Ferromagnets such as iron or cobalt have a finite magnetization, because most of their electrons' spins are oriented either with or against the magnetization axis, depending on the material. This magnetization direction persists without an outlet power, and is therefore stable[2,3]. For that reason, spintronic applications based on metallic ferromagnetic nanostructures[3] — such as magnetic hard drives and, more recently, magnetic random access memories (MRAMs) — have already proved commercial hits. But for other applications, such as reprogrammable logic, spintronics has not yet broken through into the industrial mainstream.

For that to happen, spintronics must conquer silicon, the abundant, inexpensive and entrenched material of choice for conventional semiconductor electronics. The spin of silicon's electrons is believed to survive sufficiently long to allow the persistence of spin-encoded information, and silicon-based devices might offer significant improvements on proposed spin transistors and spin-based quantum-computation schemes[2,4–7]. Yet a demonstration of even basic spintronic ingredients, such as spin injection, transport, manipulation and detection, has been elusive in silicon[2]. So why has silicon resisted for so long, when other semiconductors, such as the gallium arsenide (GaAs) used in mobile-phone electronics, have proved more pliant?

The two established ways of introducing spin into semiconductor materials are optical and electrical spin injection[2]. In optical injection, a semiconductor absorbs circularly polarized light, generating, through transfer of angular

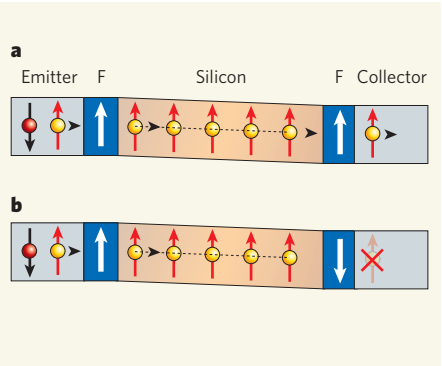


Figure 1 | Spin-valve effect. A ferromagnet (F), shown here as the perfect spin filter, causes electrons of the opposing spin orientation to bounce around and deprives them of energy, while letting electrons of the same orientation smoothly through, rather as a player in a football or hockey match deals with a member of the opposition or of his own team. As Appelbaum and colleagues demonstrate[1], putting such a filter on either side of a silicon substrate allows not only spin injection, but also spin detection. If both filters have the same orientation (**a**), electrons of that orientation will pass through the silicon from emitter to collector. If, however, the two orientations are antiparallel (**b**), electrons of neither spin will pass both filters. The relative difference in the collector current in these two cases is the spin-valve signal[3,9].

momentum, electron spin. The reverse process, the emission of circularly polarized light, is often used as smoking-gun proof of spin-injection. Unfortunately, the optical method, which works well with GaAs, is ineffective in silicon owing to subtleties of the element's electronic structure[4,8].

In electrical injection, a voltage drop typically drives spin-polarized electrons from a ferromagnetic electrode into the semiconductor. But difficulties associated with interfaces between ferromagnets and silicon, as well as a lack of a reliable method for spin detection, have hindered conclusive evidence for spin injection and transport in silicon[2,4].

Until, that is, Appelbaum *et al.* came along with their 'spin valve'[1,3,9]. In this device (Fig. 1), initially unpolarized electrons (with an equal mixture of spins pointing 'up' and spins pointing 'down') pass through a ferromagnet. This acts as a spin filter: it lets spins of only one orientation through to the silicon substrate. A second ferromagnet at the other end of the substrate will, according to whether its magnetization is the same as, or opposite to, that of the first, filter out either none or all of the remaining electrons. With this set-up, Appelbaum and colleagues observe a change in current with magnetization that provides direct evidence for the injection and transport of spin-polarized electrons in silicon.

Appelbaum and colleagues also provide[1] more subtle, but also more convincing, proof by exploiting the rotation, or precession, of spin. This precession is brought about by a magnetic field applied to the silicon substrate at right angles to the direction of the injected spins[2,10] (Fig. 2). If the magnetizations of the two spin filters are parallel, the current passing through the device will be much smaller if the electron spins precess by 180° in their passage through the silicon than if the rotation is by, say, 0° or 360°. The angle of precession is determined by the value of the applied electric and magnetic fields, but the current will exhibit peaks (and valleys) whenever the average spin arriving at the collector is parallel (or antiparallel) to the spin allowed by the second filter.

At very large magnetic fields, the spin-precession signal should disappear altogether, owing to the Hanle effect[2,10] (Fig. 2d): random scattering of the electrons on lattice imperfections causes electron transit times to differ. At high fields, this translates into spin-precession angles that differ in proportion to the transit times. If the spread in angles reaches 360°, electrons arriving at the second filter have spins pointing in all directions, and zero average spin. The collector current will cease to exhibit peaks and valleys, having instead a rather featureless dependence on the magnetic field. Remarkably, Appelbaum and colleagues observe[1] all these predicted behaviours.

Now that this proof of concept has been established, what remains to be done? First, we should find ways to raise the operating temperature from Appelbaum and colleagues' 85 kelvin to room temperature, and to increase the spin-valve signal, presently just 2%. A likely culprit for this small signal is that, because the ferromagnetic metal is grown directly on the silicon, a non-magnetic interface layer forms that is detrimental to spin transport. A similar scheme in GaAs gives a signal orders of magnitude higher[9]. There, spin injection has been achieved by using high-quality interfaces with a ferromagnetic semiconductor[2,4], or by inserting oxide tunnel barriers[11,12] between a ferromagnetic metal and the semiconductor. In silicon, low-resistance tunnel barriers next to the ferromagnets[13,14] might also help to increase the signal.

Second, whereas Appelbaum *et al.* use pure silicon, commercial electronics, as well as proposals for spintronic devices[2], relies on making silicon impure by the process known as doping. This doping creates extra charge carriers (electrons or holes), and the effect this will have on spin injection and spin precession must be investigated.

Finally, spin injection in silicon does not automatically mean that useful devices can be fabricated straight off — further ideas to explore the possibilities of silicon spin within realistic device settings are needed. Although we might not yet be quite ready for commercial silicon-based spintronics, nothing to stop the show is in sight. ■

Igor Žutić is in the Department of Physics, State University of New York at Buffalo, Buffalo, New York 14260, USA. Jaroslav Fabian is at the Institute of Theoretical Physics, University of Regensburg, 93040 Regensburg, Germany.
e-mails: zigor@buffalo.edu; jaroslav.fabian@physik.uni-regensburg.de

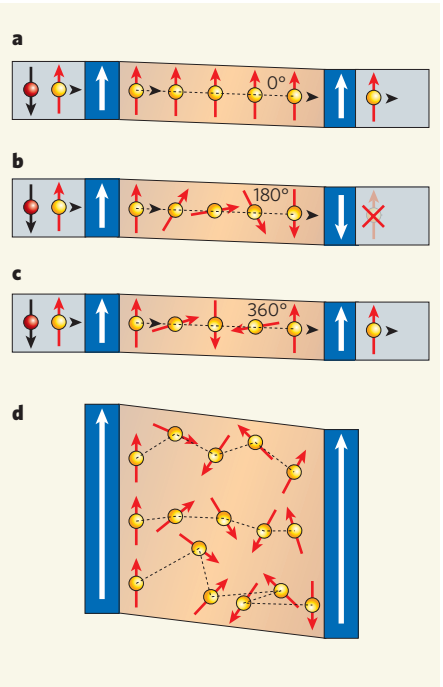


Figure 2 | Spin precession. In a magnetic field (applied perpendicular to the page here), electron spins precess with a frequency proportional to the field strength, giving rise to an oscillatory current. **a**, No precession, as in Figure 1: given two filters of the same orientation, a spin-polarized current will pass. **b**, 180° precession: no current can pass. **c**, 360° precession: a spin-polarized current passes again. **d**, If the applied field is very strong, the spins precess very fast, and the random or diffusive component of electrons' motion as they scatter off lattice imperfections comes to dominate. The angles of precession vary wildly, and spin precession is no longer detected.

1. Appelbaum, I., Huang, B. & Monsma, D. J. *Nature* **447,** 295–298 (2007).
2. Žutić, I., Fabian, J. & Das Sarma, S. *Rev. Mod. Phys.* **76,** 323–410 (2004).
3. Maekawa, S. & Shinjo, T. (eds) *Spin Dependent Transport in Magnetic Nanostructures* (Taylor & Francis, New York, 2002).
4. Žutić, I., Fabian, J. & Erwin S. C. *Phys. Rev. Lett.* **97,** 026602 (2006).
5. Sugahara, S. *IEE Proc. Circuit Devices Syst.* **152,** 355–365 (2005).
6. Tyryshkin, A. M., Lyon, S. A., Jantsch, W. & Schäffler, F. *Phys. Rev. Lett.* **94,** 126802 (2005).
7. Das Sarma, S., de Sousa, R., Hu, X. & Koiller, B. *Solid State Commun.* **133,** 737–746 (2005).
8. Lampel, G. *Phys. Rev. Lett.* **20,** 491–493 (1968).
9. Parkin, S. S. P. *et al. Proc. IEEE* **91,** 661–680 (2003).
10. Johnson, M. & Silsbee, R. H. *Phys. Rev. Lett.* **55,** 1790–1793 (1985).
11. Hanbicki, A. T. *et al. Appl. Phys. Lett.* **80,** 1240–1242 (2002).
12. Jiang, X. *et al. Phys. Rev. Lett.* **94,** 056601 (2005).
13. Min, B.-C., Motohashi, K., Lodder, C. & Jansen, R. *Nature Mater.* **5,** 817–822 (2006).
14. Žutić, I. *Nature Mater.* **5,** 771–772 (2006).

NEUROSCIENCE

Wrestling with SUMO

Françoise Coussen and Daniel Choquet

The process of SUMOylation affects various cellular events by modifying the proteins involved. In neurons, it controls receptor numbers on the cell surface, thereby regulating neuronal communication.

With the identification of cellular mechanisms that control the efficiency of the dialogue between neurons, efforts to understand how the brain adapts to new stimuli have concentrated on molecular aspects of neuronal function. For example, the internalization and intracellular trafficking of receptors for neurotransmitters — chemicals that mediate communication between neurons — has emerged as a way of regulating receptor numbers on the neuronal surface and at synapses, the junctions between neurons. This, in turn, allows reduced neuronal transmission during synaptic plasticity (the ability of synapses to change in strength), through a series of cellular processes thought to underlie learning and memory.

Internalization of cell-surface receptors and their transport to intracellular compartments are mediated by endocytosis, a process in which patches of the cell membrane are pinched off to form small vesicles carrying the cargo. On page 321 of this issue, Martin and colleagues[1] describe the specific modification of kainate receptors by a process known as SUMOylation, which triggers their endocytosis and consequently regulates synaptic function*.

Glutamate is the main excitatory neurotransmitter in neurons and binds to both ion-channel (ionotropic) and non-ion-channel (metabotropic) receptors. The three main subtypes of ionotropic glutamate receptor are kainate receptors (KARs)[2], AMPA receptors and NMDA receptors, which are also activated by the chemical compounds kainate, AMPA and NMDA, respectively. KARs are involved in the regulation of synaptic transmission and neuronal excitability by acting at pre-, post- and extrasynaptic sites; the physiological function of KARs depends on their specific localization and density at these three sites.

KARs are composed of several protein subunits, and these show great variation in the carboxy-terminal structural domains that lie on the inner, cytoplasmic, side of the neuron, and which are used for interactions with intracellular partners. The GluR6 subunit of KARs is involved in mediating the effects of kainate on inducing epilepsy[3], and is expressed at both pre- and postsynaptic sites throughout the brain, being especially enriched in the brain region known as the hippocampus.

Martin and Henley had previously demonstrated[4] that kainate induces KAR internalization, and that this depends on a regulatory enzyme called PKC, and leads to the degradation of KARs. They now[1] show that, in postsynaptic neurons, the GluR6 subunit of KARs binds to a small protein called SUMO (small ubiquitin-related modifier), which can link covalently to a lysine amino acid on its target protein. SUMOylation occurs in response to kainate or glutamate, and SUMOylated GluR6 is rapidly internalized through endocytosis, leading to a decrease in KAR-dependent synaptic activity (Fig. 1). Conversely, inhibition of the SUMO pathway increases KAR-dependent synaptic transmission. SUMOylation therefore permits a rapid and reversible modification of GluR6, altering its membrane localization and function.

SUMOylation has been most extensively studied in the context of nuclear proteins[5]. However, new roles for SUMOs have emerged, particularly in excitable cells[6] such as neurons and skeletal-muscle cells, where SUMOs regulate the function of cell-membrane proteins. The SUMO-conjugation machinery is found throughout these cells and is present at synapses. The K2P1 and Kv1.5 K^+ ion channels, glucose transporters and group 3 metabotropic glutamate receptors are all SUMOylation targets. For example, SUMOylation of the K2P1 channel prevents its opening[6], but has a modulatory effect on the Kv1.5 channel[7]. Thus, reversible control of ion channels and receptors through this process may be a widespread mechanism for tuning the electrical excitability of cells. Martin *et al.*[1] extend this concept to ionotropic glutamate receptors.

A parallel can be drawn between SUMOylation and ubiquitination, which also occurs after protein translation and targets proteins for degradation. Internalization of GluR6 is also promoted by ubiquitination[8] (Fig. 1). GluR6 target sites for ubiquitination and SUMOylation are only three amino acids apart. So, could SUMO and ubiquitin be added simultaneously to GluR6? What is the interplay between these two protein-modifying mechanisms? As GluR6 is localized in both pre- and postsynaptic sites, could the two processes act in different subcellular compartments? SUMOylation seems to be specific for KARs. By contrast, ubiquitination also mediates the degradation of AMPA and NMDA receptors, which are involved in fast synaptic transmission[9], although possibly by different mechanisms[10]. On the other hand, GluR6 is probably a constituent subunit of most tetrameric KARs, and its SUMOylation could mediate the endocytosis of heterotetramers.

Many other questions also remain. First, for example, how is the SUMOylation of KARs triggered? There are several possibilities. Binding of kainate to KARs could lead to a conformational change allowing its

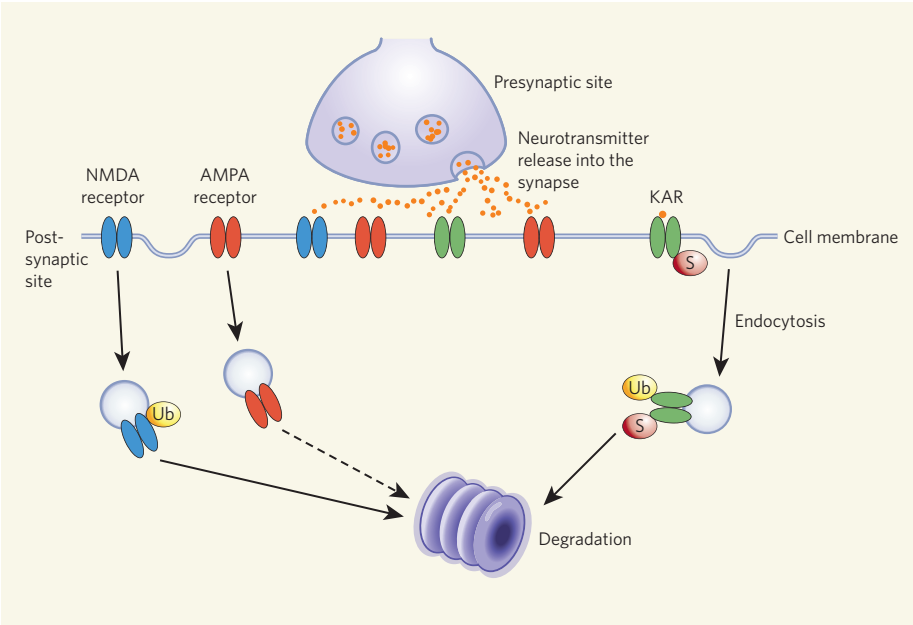


Figure 1 | Post-translational modification and the regulation of ionotropic glutamate receptors. Glutamate receptors, such as kainate receptors (KARs) and the AMPA and NMDA receptors, diffuse in and out of synapses, and are removed from the neuronal surface by endocytosis and degraded. Martin *et al.*[1] show that direct SUMOylation (S) of the GluR6 subunit of KARs induces their endocytosis in neuronal cells on activation by glutamate or kainate. Ubiquitination (Ub), another form of post-translational modification, which targets proteins for degradation, regulates surface expression of all ionotropic glutamate receptors, albeit probably by different mechanisms.

*This article and the paper concerned[1] were published online on 7 May 2007.

SUMOylation. Alternatively, kainate could induce a specific signal-transduction pathway to trigger SUMOylation, as the involvement of PKC in kainate-induced KAR endocytosis might suggest[4]. Finally, the carboxy-terminal domain of GluR6 binds to scaffolding proteins such as Pick1 and syntenin. SUMOylation could interfere with these interactions, thus promoting KAR endocytosis by dislodging it from its anchors at the cell surface and/or enhancing its interaction with members of the endocytic pathway.

What about functional consequences? SUMOylation-mediated endocytosis of synaptic KARs is known to occur under physiological conditions, because the SUMO inhibitor SENP1 induces an increase in spontaneous miniature excitatory postsynaptic currents. So, could the SUMOylation of KARs be involved in synaptic plasticity? Alternatively, given that KARs are involved in neuronal death due to receptor overactivation (excitotoxicity)[3], Martin *et al.*[1] propose that SUMOylation-induced endocytosis could help protect neurons from excitotoxic stress.

The findings of Martin *et al.* add SUMOylation to the list of post-translational molecular mechanisms that regulate glutamate-receptor trafficking, such as phosphorylation and ubiquitination, and indicate its possible involvement in synaptic plasticity. The challenge is to sort out the interplay between these mechanisms, and to unravel the precise physiological context in which they control receptor numbers and function. ■

Françoise Coussen and Daniel Choquet are at the Institut François Magendie, Physiologie Cellulaire de la Synapse, UMR CNRS 5091, Université de Bordeaux, Bordeaux 33077, France.
e-mails: fcoussen@u-bordeaux2.fr; dchoquet@u-bordeaux2.fr

1. Martin, S., Nishimune, A., Mellor, J. R. & Henley, J. M. *Nature* **447,** 321–325 (2007).
2. Pinheiro, P. & Mulle, C. *Cell Tissue Res.* **326,** 457–482 (2006).
3. Mulle, C. *et al. Nature* **392,** 601–605 (1998).
4. Martin, S. & Henley, J. M. *EMBO J.* **23,** 4749–4759 (2004).
5. Seeler, J. S. & Dejean, A. *Nature Rev. Mol. Cell Biol.* **4,** 690–699 (2003).
6. Scheschonka, A., Tang, Z. & Betz, H. *Trends Neurosci.* **30,** 85–91 (2007).
7. Benson, M. D. *et al. Proc. Natl Acad. Sci. USA* **104,** 1805–1810 (2007).
8. Salinas, G. D. *et al. J. Biol. Chem.* **281,** 40164–40173 (2006).
9. Bingol, B. & Schuman, E. M. *Curr. Opin. Neurobiol.* **15,** 536–541 (2005).
10. Yi, J. J. & Ehlers, M. D. *Pharmacol. Rev.* **59,** 14–39 (2007).

ORGANIC CHEMISTRY

Molecular cross-talk

Alexander Greer

There is a long way to go before artificial enzymes can reproduce the functions of the real things. The advent of systems that generate and respond to signals may bring that ideal a step closer.

Molecules are often used as signals in biological systems, triggering reactions to various stimuli. Enzymes in plants, for example, liberate electronically excited oxygen — known as singlet oxygen — in response to stress, eliciting protective responses from other biomolecules[1,2]. Such chemical cross-talk has now been demonstrated in a synthetic system, as reported by Natarajan *et al.*[3] in the *Journal of the American Chemical Society*. They have designed an artificial enzyme that generates singlet oxygen and releases it into the surrounding solution. The oxygen then diffuses to another, remote artificial enzyme, where it is trapped and reacts with acceptor molecules.

Artificial enzymes have been around since the 1980s, when chemists discovered that the hydrophobic interior environment of enzymes could be mimicked by large 'host' molecules that encapsulate small 'guest' molecules[4]. But despite the great ingenuity applied to the design of the hosts, these systems were primitive compared with natural, highly evolved enzyme pockets. One big problem was devising recognition and capturing strategies for molecules of sufficient quality to accurately direct guests to the artificial-enzyme sites — such exquisite control is essential for reproducing the functions of natural enzymes. An understanding of molecular host–guest binding has slowly developed so that, for example, we now appreciate how non-covalent interactions can immobilize guests in a process known as constrictive binding[5]. It is also possible to tailor the size and shape of hosts, so that their geometry can direct chemical reactions to selected sites in a guest molecule[6,7].

Natarajan *et al.* have used such ideas in their study[3]. They trapped two different guest molecules into separate hosts, permanently isolating them from each other (Fig. 1, overleaf). The first guest was a photosensitizer — a compound that readily transfers light energy to other molecules, so promoting them into excited states. The second isolated guest was an alkene, which acts as an oxygen acceptor. The encapsulated photosensitizer absorbs light and — if the host is open to allow oxygen into its cavity — converts ground-state oxygen into singlet oxygen, which then escapes into the bulk solution. Some of the singlet oxygen reaches the separate alkene-containing hosts, where it reacts with the guest molecule. Because the alkene-containing hosts hold their guests in a particular orientation within the cavity, some parts of the alkene are more accessible to chemical attack than others. The singlet oxygen therefore predominantly attacks the alkene at the most accessible position, yielding mostly one product. In the absence of the host, the singlet oxygen attacks the alkene at several positions, so that a mixture of compounds forms, with no major product[8].

Despite the effectiveness of this reaction system, surprisingly little is known about the dynamics of the artificial enzymes. For example, how does the oxygen enter the cavities? The hosts are actually dimeric complexes, in which two bowl-shaped subunits are held together by weak intermolecular forces to form capsules[9]. The structural integrity of these capsules is therefore not fixed, because they are not rigidly bound together by covalent bonds. The authors suggest[3] that the alkene-containing host probably opens up to allow singlet oxygen in, rather than the oxygen somehow penetrating the walls of the closed capsule (which would be a much slower process than that observed).

Some details of the opening and closing process may be gleaned from previous studies on capsules in which the two halves are held together with covalent bonds[5]. These indicate that the gating mechanism is akin to the movement of French doors — where a couple of small flaps hinge apart to create an opening — rather than a more severe deformation that prises apart the entire host. This is reminiscent of gating in certain natural enzymes, such as HIV-1 protease, in which a flap over the active site can flip-flop for the exclusion or inclusion of guests[10,11]. Future work in this area would undoubtedly benefit from a computational analysis of how the structure swings open. An assessment of the attack trajectories of the singlet oxygen would also help in the design of other artificial enzymes that attempt to control the exact position of a reaction in a molecule[12].

Singlet oxygen is highly reactive, and is rapidly quenched in water to form stable, ground-state oxygen. To maximize the amount of singlet oxygen that reaches the target sites in Natarajan and colleagues' system[3], two approaches can be envisaged: the oxygen molecules could be stabilized in some way that enables them to diffuse a greater distance before being quenched, or protective channels could be provided between hosts to prevent quenching. But quenching reactions might actually be useful, as they can steer the attack of singlet oxygen onto substrates, so directing the orientation of chemical groups in the reaction product[13]; singlet oxygen might quench on one side of a molecule, but react chemically on the other. Incorporating such control

50 YEARS AGO

Two sharply divergent points of view are held with regard to children's films. The Soviet bloc and most of Europe believe that children at the cinema should be sheltered from the actualities of life. They should be provided with cartoon or puppet films which show them a fairy-tale world of fantasy or else with films with a direct moral purpose such as "stressing the value of human labour". On the other hand, the Americans, who do not make special children's films, believe that children are essentially little adults and are perfectly ready to take adult screen entertainment. Neither of these divergent points of view is accepted in Britain, where it is believed that, to get the greatest pleasure and profit in the cinema, children should see specially produced films within their understanding and experience, and these should be mainly realistic... Children are really interested only in children like themselves or in attractive animals; they have little interest in adults and few cowboys would be popular if deprived of their horses. However, very old people who are approaching their second childhood compel attention.

From *Nature* 18 May 1957.

100 YEARS AGO

The problem of establishing a connection between the mineral ingredients of the tea plant and the quality and strength of the tea is under investigation, with a prospect of obtaining definite results. From a study of the methods of preparing Oolong tea in Formosa, it is concluded that the quality and characteristics are due to an aroma produced by faint oxidation in drying and a slight scorching during roasting of the leaf, as well as to the mild decomposition caused by a fungus, and it is suggested that the fungus acting on the legumin in the leaf produces flavouring bodies similar to the action of moulds in cheese.

From *Nature* 16 May 1907.

50 & 100 YEARS AGO

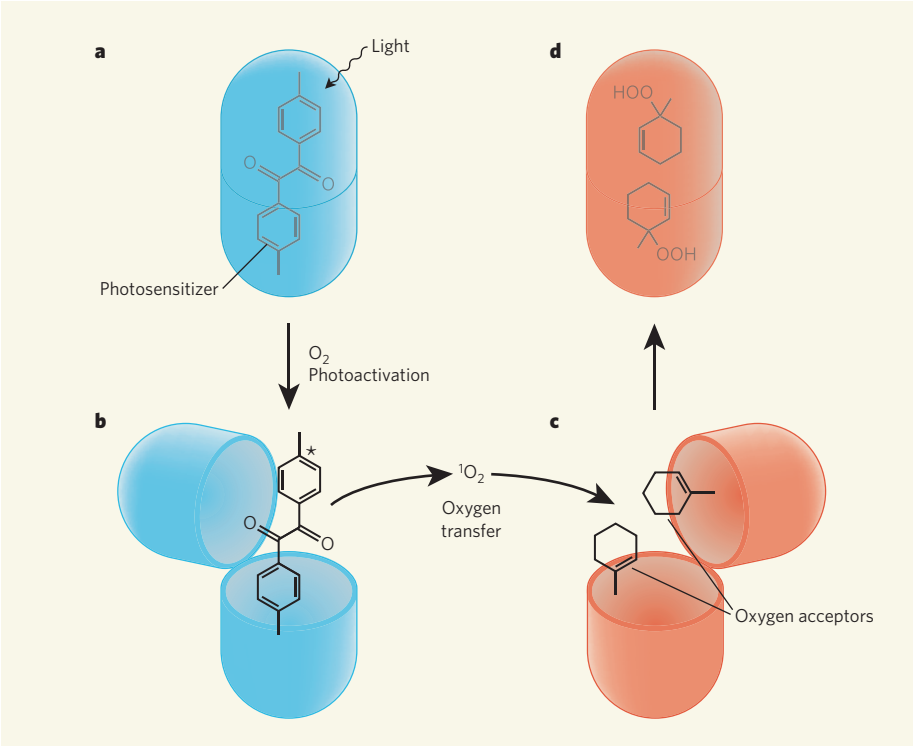


Figure 1 | Chemical correspondence. Natarajan *et al.*[3] have devised an artificial system that mimics chemical signalling between enzymes. **a**, A photosensitizer molecule is trapped in an artificial enzyme pocket and irradiated with light. **b**, The light energy promotes the photosensitizer into an excited electronic state (indicated by the asterisk), which in turn excites oxygen molecules if the enzyme pocket is open. **c**, The excited oxygen, $^1O_2$, diffuses through the surrounding solution until it reaches an open pocket of another enzyme that contains oxygen acceptor molecules. **d**, Because the acceptor molecules bind into the enzyme pocket in a particular orientation, the excited oxygen reacts only at the most accessible site of those molecules.

into enzyme-mimics that use singlet oxygen might help them target specific sites in their substrates.

Natarajan and colleagues' report[3] forms part of a growing body of work examining chemical cross-talk. Their study is carried out entirely in solution, but an interesting related area examines the transportation of molecules between solid materials, which may be more relevant to the mechanisms used by proteins for trapping molecules and responding to chemical signals. Singlet oxygen can diffuse in polymer films[14] or be taken up by acceptor molecules connected to solid supports (such as resins, or porous inorganic materials known as zeolites). Resin-to-resin reactions are quite useful because they can transfer a variety of chemical signals through solution[15], including inherently unstable molecules such as singlet oxygen and cyclobutadiene (a reactive hydrocarbon). This challenges the preconception that only stable molecules can act as chemical signals.

There is no denying that synthetic enzymes will require huge improvements if they are to compete with their biological equivalents. Problems still awaiting exploration include the incorporation of allosteric features — binding sites other than the main cavity — into artificial hosts, and the introduction of cooperative dialogue between hosts that might enhance chemical trapping, as occurs in nature. But as a first step in the advanced development of artificial enzymes, Natarajan and colleagues' work[3] on chemical signalling will certainly create a lot of cross-talk between chemists. ■

Alexander Greer is in the Department of Chemistry, Graduate Center and the City University of New York, Brooklyn College, 2900 Bedford Avenue, Brooklyn, New York 11210, USA.
e-mail: agreer@brooklyn.cuny.edu

1. Laloi, C. *et al. Proc. Natl Acad. Sci. USA* **104,** 672–677 (2007).
2. Klotz, L.-O., Briviba, K. & Sies, H. in *Antioxidant and Redox Regulation of Genes* (eds Sen, C. K., Sies, H. & Baeuerle, P. A.) 3–20 (Academic, San Diego, 2000).
3. Natarajan, A. *et al. J. Am. Chem. Soc.* **129,** 4132–4133 (2007).
4. Cram, D. J. *Angew. Chem. Int. Edn Engl.* **27,** 1009–1112 (1988).
5. Houk, K. N., Nakamura, K., Sheu, C. & Keating, A. E. *Science* **273,** 627–629 (1997).
6. MacGillivray, L. R. & Atwood, J. L. *Nature* **389,** 469–472 (1997).
7. Liu, X. & Warmuth, R. *J. Am. Chem. Soc.* **128,** 14120–14127 (2006).
8. Wasserman, H. H. & Murray, R. W. (eds) *Singlet Oxygen* (Academic, New York, 1979).
9. Conn, M. M. & Rebek, J. Jr *Chem. Rev.* **97,** 1647–1668 (1997).
10. Hornak, V., Okur, A., Rizzo, R. C. & Simmerling, C. *Proc. Natl Acad. Sci. USA* **103,** 915–920 (2006).
11. Tozzini, V., Trylska, J., Chang, C.-E. & McCammon, J. A. *J. Struct. Biol.* **157,** 606–615 (2007).
12. Carpenter, B. K. in *Reactive Intermediate Chemistry* (eds Moss, R. A., Platz, M. S. & Jones, M. Jr) 925–960 (Wiley, Hoboken, 2004).
13. Sivaguru, J. *et al. Tetrahedron* **62,** 10647–10659 (2006).
14. Gao, Y. & Ogilby, P. R. *Macromolecules* **25,** 4962–4966 (1992).
15. Rebek, J. Jr *J. Org. Chem.* **69,** 2651–2660 (2004).

PLANT DEVELOPMENT

Parental conflict overcome

Nir Ohad

In flowering plants, viable seeds result even without two of the mechanisms that normally operate during embryogenesis. This finding illuminates the interplay of male and female factors in the process.

In flowering plants, two identical male gametes from the same pollen grain fuse in the ovule with two female gametes, initiating development of the embryo and the endosperm — the tissue that nourishes the embryo, equivalent to the placenta in mammals. As they describe on page 312 of this issue, Nowack *et al.*[1] find that when only one male gamete triggers embryogenesis, a viable seed can develop if the maternal mechanism controlling endosperm development is disabled. This finding reveals a new layer of conflicting interaction between the maternal and paternal genomes in regulating plant embryogenesis, and supports an early theory concerning endosperm evolution.

Whereas a single fertilization event initiates embryogenesis in animals, in flowering plants two genetically identical gametes from the same pollen grain fertilize the haploid egg cell (chromosome complement n) and the diploid ($2n$) central cell within the ovule. The result is a diploid ($2n$) embryo and a triploid ($3n$) endosperm (Fig. 1a). *In vitro* systems have been used to study the double-fertilization process, but no such *in vivo* studies have been successful so far.

Nowack and co-workers, however, have dissected the regulatory mechanism controlling endosperm and embryo development *in vivo*, and show that seeds with uniparental endosperm of maternal origin are viable. This was achieved using a novel mutant of a model plant, *Arabidopsis thaliana*, that is impaired in the *CDKA;1* gene. Instead of the normal two male gametes, only one gamete develops in the *cdka;1* mutant pollen, which exclusively fertilizes the egg cell[2]. Using the *cdka;1* mutant as a pollen donor, the authors previously showed that this fertilization of the egg cell induces the development of a diploid embryo[2]. And even though the second fertilization event did not take place, endosperm developed. However, this 'homodiploid' ($2n$) endosperm was abnormal and the seeds aborted (Fig. 1b). Endosperm development in the absence of the second fertilization event led the authors[2] to conclude that the fertilized egg cell may provide a signal that can promote the initial stages of this process, thus overcoming the maternal repression within the central cell. Normally, the second fertilization event overcomes this repression.

Such maternal repressive mechanisms act epigenetically. That is, they involve DNA methylation[3] and modification of the histone proteins associated with DNA[4–6], a process partly mediated by complexes of the Polycomb group proteins (PcG)[3–5,7]. Mutations in maternal PcG members (*fie*[8], *fis2*[9] or *mea*[10]) also lead to the proliferation of abnormal ($2n$) endosperm in the absence of fertilization (Fig. 1c). If fertilization does take place in these mutants, normal diploid embryo and abnormal triploid endosperm develop, leading to seed abortion.

Nowack and co-workers[1] have taken dissection of the double-fertilization process a step further, addressing the question of how far the embryo can develop in the absence of the second fertilization event. When the *cdka;1* mutant was used as a pollen donor to fertilize female mutant plants impaired in *fie*, *fis2* or *mea*, homomaternal diploid ($2n$) endosperm developed. Surprisingly, this endosperm developed normally, giving rise to a seed that could germinate and produce a viable plant (although of reduced stature) (Fig. 1d). These results show that even when only the egg cell is fertilized, seed development can be completed if maternal repression of central-cell division is bypassed. This situation occurs when the paternal genome does not contribute genetic input to the endosperm and when the maternal PcG machinery restricting endosperm proliferation is impaired. Remarkably, under such

GENETICS

Run, whippet, run

Whippets are fast — these dogs can run at speeds of more than 60 kilometres per hour, and have long been used in racing. But, recently, owners have been reporting an increase in 'double muscling', a trait characterized by cramping in the shoulders and thighs of otherwise healthy whippets. Dana Mosher and colleagues now show that a mutation in the *MSTN* gene is the culprit (D. S. Mosher *et al. PLoS Genet.* e79.eor doi:10.1371/journal.pgen.0030079.eor; 2007).

The *MSTN* gene encodes myostatin, a protein that affects muscle composition. Mosher *et al.* analysed the sequence of *MSTN* from three categories of whippet: the heavy-muscled 'bully' whippets, which have double muscling; parents of bully whippets; and normal whippets unrelated to bullies.

They found that all bully whippets carry a mutation in both copies of the *MSTN* gene, and all bully-whippet parents have the same mutation in one copy of this gene. This mutation involves deletion of only two DNA base pairs, resulting in premature termination of messenger RNA translation and a truncated form of myostatin.

Looking for a correlation between this *MSTN* mutation and whippets' racing performance, the authors observed a significant association between the dogs' speed, their muscle mass and their genetic profile. Whippets with one mutated copy of *MSTN* were generally the fastest, followed by normal animals. Bully whippets are rarely raced, probably because they are handicapped by an excess of muscle. The increase in the incidence of the *MSTN* mutation is probably due to selective breeding of whippets.

BLICKWINKEL/ALAMY

Mosher *et al.* did not detect this mutation in 14 other heavy-muscled breeds of dog, suggesting that it might be unique to whippets. There is the question of whether a similar mutation in humans could be exploited to enhance athletic performance. However, the authors caution that effects of the *MSTN* mutation on other aspects of health are unknown.

Sadaf Shadan

NATURE|Vol 447|17 May 2007

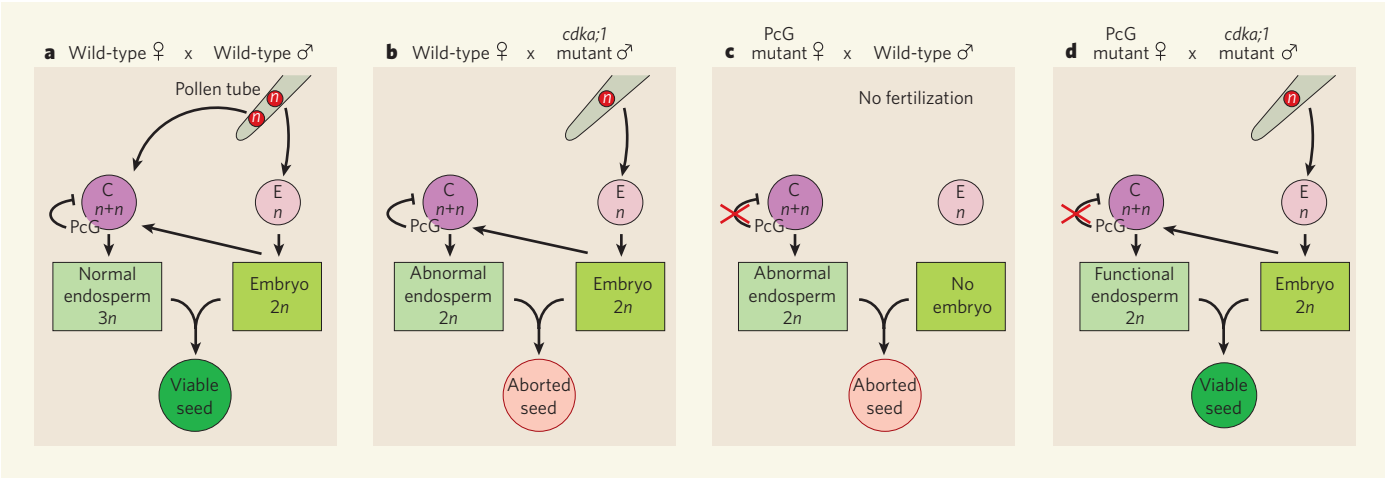


Figure 1 | Plant fertilization and seed development. **a**, In a cross between wild-type (non-mutant) female and male, the pollen tube contains two identical gametes (red circles): one fertilizes the egg cell (E) and the other fertilizes the central cell (C) in the ovule. Embryo and endosperm then develop into a viable seed. **b**, Pollen tubes of *cdka;1* mutants contain only one gamete, which fertilizes the egg cell only. The embryo prompts development of the endosperm, but abnormally so, and the result is aborted seed[2]. **c**, In mutants in which genes encoding members of the Polycomb group (PcG)[8–10] are affected, the PcG machinery does not inhibit the mature central cell from proliferating in the absence of fertilization. In this case, homomaternal (2*n*) abnormal endosperm develops, and the seed aborts. **d**, As Nowack *et al.*[1] now show, development of small but viable seeds with unimaternal diploid endosperm can occur when the maternal imprinting is bypassed and only the egg cell is fertilized.

conditions, maternally derived endosperm displays the characteristics of normal endosperm and is sufficient to sustain embryo development.

These findings point to conflicting interactions between the parental genomes. They are in accordance with the kinship theory[11], which predicts that, when sexual reproduction occurs, each parent will promote the interests of its own genes. This is achieved by the selection of mechanisms that control the maternal nutrition of the embryo. The parental conflict results in genomic 'imprinting'[3] of genes essential for embryo nutrition, leading to different gene-expression profiles imposed by each of the parental genomes.

The observations[1] give definitive backing to a hypothesis, proposed in 1900 by Eduard Strasburger[12], which ascribed a maternal origin to the endosperm that develops once the egg cell is fertilized. The fact that in non-seeded plants (gymnosperms), which lack double fertilization, the endosperm-equivalent nourishing tissue is of maternal origin also supports the hypothesis.

The findings of Nowack *et al.*[1] further point to the possibility that, during the evolution of the double-fertilization system in flowering plants, existing PcG machinery was mobilized to participate in seed development (interestingly, in mammals, PcG controls placenta development, among other functions[7]). The new study also provides an entry point for discovering the signalling molecules that coordinate the cross-talk between the developing embryo and endosperm. Moreover, the genetic tools developed in this research will help in identifying the downstream components activated during either embryo or endosperm development. Other imprinting mechanisms, such as DNA methylation[3], contribute to plant embryogenesis, and similar genetic tools could be used to investigate them.

A question remains: if both the double-fertilization and maternal regulatory systems can be removed without affecting plant viability, why did such a complex system evolve? ■

Nir Ohad is in the Department of Plant Sciences, Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel.
e-mail: niro@tauex.tau.ac.il

1. Nowack, M. K. *et al. Nature* **447,** 312–315 (2007).
2. Nowack, M. K. *et al. Nature Genet.* **38,** 63–67 (2006).
3. Gehring, M., Choi, Y. & Fischer, R. *Plant Cell* **16,** S203–S213 (2004).
4. Schubert, D., Clarenz, O. & Goodrich, J. *Curr. Opin. Plant Biol.* **8,** 553–561 (2005).
5. Kohler, C. & Makarevich, G. *EMBO Rep.* **7,** 1223–1227 (2006).
6. Calonje, M. & Sung, Z. R. *Curr. Opin. Plant Biol.* **9,** 530–537 (2006).
7. Guitton, A. & Berger, B. *Int. J. Dev. Biol.* **49,** 707–716 (2005).
8. Ohad, N. *et al. Proc. Natl Acad. Sci. USA* **93,** 5319–5324 (1996).
9. Chaudhury, A. *et al. Proc. Natl Acad. Sci. USA* **94,** 4223–4228 (1997).
10. Grossniklaus, U., Vielle-Calzada, J. P., Hoeppner, M. & Gagliano, W. *Science* **280,** 446–450 (1998).
11. Haig, D. & Westoby, M. *Am. Nat.* **134,** 147–155 (1989).
12. Strasburger, E. *Bot. Z.* **58,** 294–315 (1900).

PLANETARY SCIENCE

Cracks under stress

Andrew J. Dombard

Two modelling studies provide complementary descriptions of how gravitational forces might help to form the plumes of water vapour that spout from cracks in Enceladus, one of Saturn's icy moons.

Enceladus, an icy satellite of Saturn only about 500 kilometres in diameter, is a remarkably active body[1–3]. NASA's Saturn orbiter, Cassini, observed plumes composed mostly of water, presumed to be spouting from a set of nearly parallel cracks, nicknamed the 'tiger stripes'. These cracks run for about 130 km across the moon's southern polar region and are warmer than the surrounding terrain (Fig. 1). What exactly causes the venting is not clear, although gravitational 'tidal' forces within the Saturn system are thought to be the ultimate culprit. In this issue, Nimmo *et al.* (page 289)[4] postulate that heating from shearing motion along the tiger stripes, caused by tidal forces, provides energy for the activity. Meanwhile, Hurford *et al.* (page 292)[5] consider how forces across the stripes might open and shut the cryovolcanic vents.

Tidal forces acting on Enceladus combine with its spin to distort the moon's shape into an ellipsoid whose longest axis points towards Saturn. For a circular orbit at constant speed, these tides would be constant, and Enceladus quiescent. But Enceladus's orbit is eccentric — it deviates from a perfect circle. Enceladus is thus sometimes closer to, and sometimes farther away from, its parent planet, and the size

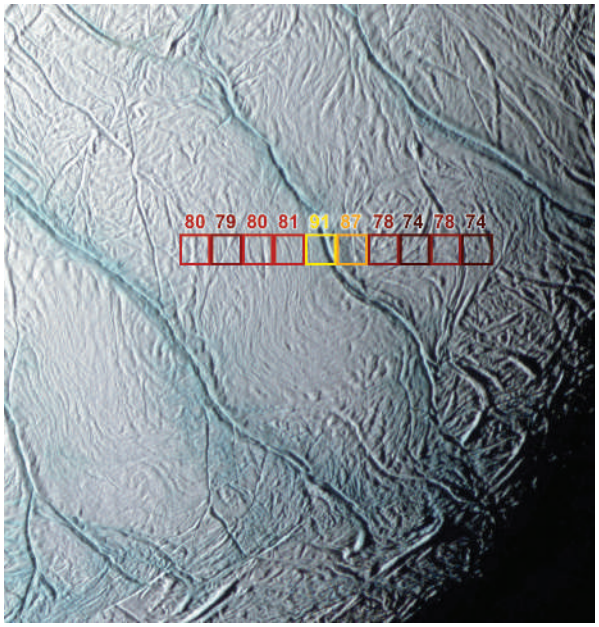
of its tidal bulge varies accordingly.

As Kepler's second law states, an eccentric orbit also results in a variable orbital velocity: Enceladus orbits faster when closer to Saturn, and more slowly when farther away. Its rotation around its own axis, on the other hand, is constant, at one revolution every 1.37 Earth days. So the rotational and orbital motions are out of phase, and the line linking Saturn and Enceladus oscillates around a fixed point on the moon's surface (Fig. 2). This results in a cyclical warping of Enceladus's shape each day that dissipates heat within it; such a mechanism also heats the interior of Io, a satellite of Jupiter and the most volcanically active body in the Solar System. But the magnitude of this heat source for Enceladus is insufficient to drive the activity around its southern pole[1].

Nimmo *et al.*[4] take a different tack, examining how heating might become concentrated at the tiger stripes. The changing size and position of the tidal bulge produces a predictable pattern of stress that cycles across the surface every orbit. At every point along the tiger stripes, these stresses can be decomposed into a shear component acting along the crack, and a normal component acting at right angles to it. The directions of both components switch from one half of the orbit to the next.

The authors propose that, rather as rubbing one's hands together on a cold day makes them warm, lateral motion along the crack provides the energy to drive the venting activity. This energy is partitioned into heating of the surrounding terrain and vapour sublimation. The best match to the observations comes when 10% of the energy goes into heating the terrain and the rest goes into vapour production, with 90% of the vapour recondensing in the shallow subsurface near the tiger stripes. In contrast to indications from past work[1], temperatures immediately beneath Enceladus's surface need not be driven up to the melting point of water in this process, although the model best matches the observations when the moon's outer ice shell is mechanically decoupled from the rocky core. This decoupling is presumed to be due to an ocean of liquid water at depth, which makes the shell malleable enough to permit sufficient heating.

Hurford *et al.*[5] examine the effect of normal stresses across the tiger stripes. They propose that tensile stresses during half of the orbit open the stripes, facilitating eruption, and compressive stresses inhibit eruption during the other half. For three close Cassini fly-bys, the authors predict whether the state of normal stress across every point along the tiger stripes was tensile during the entirety of each time frame, whether it was compressive, or whether it was transitional between the two. These predictions of how conducive the stress state is to eruption can be compared with observations, although images of the plumes acquired during these fly-bys do not lend themselves easily to such an exercise. For an additional close fly-by that occurred on 24 April 2007, Hurford *et al.* forecast[5] that a large fraction of the tiger stripe Damascus Sulcus would be either in tension or transitioning to such a state. The viewing geometry of this encounter is such that a plume issuing from Damascus will be spatially distinct, allowing a more thorough testing of the hypothesis that tidal stresses control the venting.



NASA/JPL/SPACE SCIENCE INST.

Figure 1 | Tiger stripes burning bright. Infrared spectrometer measurements of the series of almost parallel cracks — 'tiger stripes' — near Enceladus's southern pole, taken by NASA's Cassini orbiter during a fly-by on 14 July 2005, reveal their temperature to be as high as 91 kelvin (–296 °C). By contrast, the surrounding surface temperatures vary between 74 K and 81 K.

Nimmo *et al.*[4] and Hurford *et al.*[5] provide new directions for understanding Enceladus with complementary, but separate, investigations of the consequences of the shear and normal stresses across the tiger stripes. The influence of the unmodelled stress component is not fully considered in either paper, however. Tensile normal stress across a tiger stripe during half of the orbit might reduce the frictional contact between the crack surfaces, thus limiting the amount of shear heating. Compressive normal stress during the other half of the orbit, on the other hand, could result in sticking of the crack. This more limited slip, and potentially full locking, might produce large strike-slip motion on the tiger stripes. But such large strike-slip offset is not observed, thus calling into question how effective a barrier to eruption the compressive state is.

Also, these models[4,5] describe only how the satellite is behaving, not why. In stark contrast to Enceladus, Mimas, another icy satellite of Saturn with a closer orbit and a larger eccentricity, has followed a benign evolution. Why Enceladus has ended up in such an active state whereas Mimas has not are questions that will exercise planetary scientists for some time to come. ■

Andrew J. Dombard is in the Applied Physics Laboratory, Johns Hopkins University, 11100 Johns Hopkins Road, Laurel, Maryland 20723-6099, USA.
e-mail: andrew.dombard@jhuapl.edu

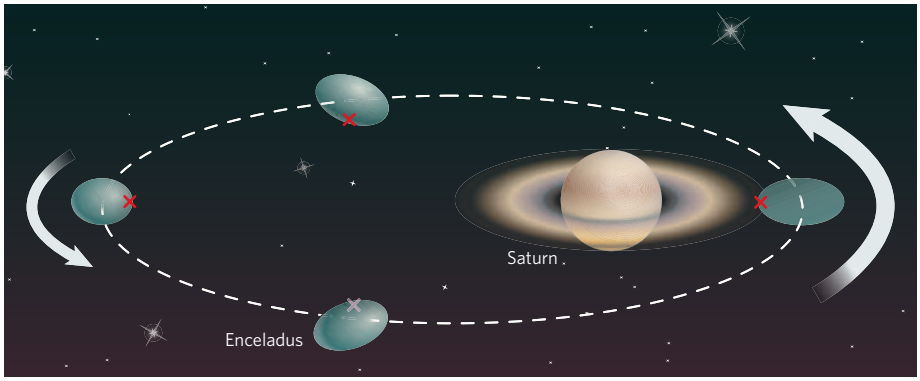


Figure 2 | Mobile bulge. Enceladus's eccentric orbit means it is sometimes nearer to, sometimes farther away from, its parent planet Saturn. The tidal bulge of Enceladus along the axis connecting the moon with Saturn accordingly varies in size (exaggerated here). Another effect of the eccentricity is that Enceladus's orbital speed is not constant: according to Kepler's second law, an orbiting body travels faster when it is closer to its parent. Enceladus's speed of rotation around its own axis, on the other hand, is constant. The orbital and rotational motions are thus out of phase, with the result that the position of Enceladus's tidal bulge oscillates around a fixed point (red crosses; pink cross indicates point on far side of planet). This constant warping of the moon's surface might cause the stress at the cracks near its southern pole that were investigated by Nimmo *et al.*[4] and Hurford *et al.*[5].

1. Porco, C. C. *et al. Science* **311,** 1393–1401 (2006).
2. Spencer, J. R. *et al. Science* **311,** 1401–1405 (2006).
3. Waite, J. H. *et al. Science* **311,** 1419–1422 (2006).
4. Nimmo, F., Spencer, J. R., Pappalardo, R. T. & Mullen, M. E. *Nature* **447,** 289–291 (2007).
5. Hurford, T. A., Helfenstein, P., Hoppa, G. V., Greenberg, R. & Bills, B. G. *Nature* **447,** 292–294 (2007).

REVIEWS

Origins of major human infectious diseases

Nathan D. Wolfe[1], Claire Panosian Dunavan[2] & Jared Diamond[3]

Many of the major human infectious diseases, including some now confined to humans and absent from animals, are 'new' ones that arose only after the origins of agriculture. Where did they come from? Why are they overwhelmingly of Old World origins? Here we show that answers to these questions are different for tropical and temperate diseases; for instance, in the relative importance of domestic animals and wild primates as sources. We identify five intermediate stages through which a pathogen exclusively infecting animals may become transformed into a pathogen exclusively infecting humans. We propose an initiative to resolve disputed origins of major diseases, and a global early warning system to monitor pathogens infecting individuals exposed to wild animals.

Human hunter/gatherer populations currently suffer, and presumably have suffered for millions of years, from infectious diseases similar or identical to diseases of other wild primate populations. However, the most important infectious diseases of modern food-producing human populations also include diseases that could have emerged only within the past 11,000 years, following the rise of agriculture[1,2]. We infer this because, as discussed below, these diseases can only be sustained in large dense human populations that did not exist anywhere in the world before agriculture. What were the sources of our major infectious diseases, including these 'new' ones? Why do so many animal pathogens, including virulent viruses like Ebola and Marburg, periodically infect human hosts but then fail to establish themselves in human populations?

A tentative earlier formulation[1] noted that major infectious diseases of temperate zones seem to have arisen overwhelmingly in the Old World (Africa, Asia and Europe), often from diseases of Old World domestic animals. Hence one goal of this article is to reappraise that conclusion in the light of studies of the past decade. Another goal is to extend the analysis to origins of tropical diseases[3]. We shall show that they also arose mainly in the Old World, but for different reasons, and mostly not from diseases of domestic animals.

These results provide a framework for addressing unanswered questions about the evolution of human infectious diseases—questions not only of practical importance to physicians, and to all the rest of us as potential victims, but also of intellectual interest to historians and evolutionary biologists. Historians increasingly recognize that infectious diseases have had major effects on the course of history; for example, on the European conquest of Native Americans and Pacific Islanders, the inability of Europeans to conquer the Old World tropics for many centuries, the failure of Napoleon's invasion of Russia, and the failure of the French attempt to complete construction of a Panama Canal[4–6]. Evolutionary biologists realize that infectious diseases, as a leading cause of human morbidity and mortality, have exerted important selective forces on our genomes[2,7].

We begin by defining five stages in the evolutionary transformation of an animal pathogen into a specialized pathogen of humans, and by considering why so many pathogens fail to make the transition from one stage to the next. We then assemble a database of 15 temperate and 10 tropical diseases of high evolutionary and/or historical impact, and we compare their characteristics and origins. Our concluding section lays out some unresolved questions and suggests two expanded research priorities. We restrict our discussion to unicellular microbial pathogens. We exclude macroparasites (in the sense of ref. 7), as well as normally benign commensals that cause serious illness only in weakened hosts. The extensive Supplementary Information provides details and references on our 25 diseases, robustness tests of our conclusions, factors affecting transitions between disease stages, and modern practices altering the risk of emergence of new diseases.

Evolutionary stages

Box 1 delineates five intergrading stages (Fig. 1) through which a pathogen exclusively infecting animals (Stage 1) may become transformed into a pathogen exclusively infecting humans (Stage 5). Supplementary Table S1 assigns each of the 25 major diseases discussed (Supplementary Note S1) to one of these five stages.

A large literature discusses the conditions required for a Stage 5 epidemic to persist[2,7]. Briefly, if the disease infects only humans and lacks an animal or environmental reservoir, each infected human introduced into a large population of susceptible individuals must on average give rise during his/her contagious lifespan to an infection in at least one other individual. Persistence depends on factors such as the duration of a host's infectivity; the rate of infection of new hosts; rate of development of host protective immunity; and host population density, size and structure permitting the pathogen's regional persistence despite temporary local extinctions.

Less well understood are two of the critical transitions between stages, discussed in Box 2. One is the transition from Stage 1 to Stage 2, when a pathogen initially confined to animals first infects humans. The other is the transition from Stage 2 to Stages 3 and 4 (see also Supplementary Note S2), when a pathogen of animal origin that is nevertheless transmissible to humans evolves the ability to sustain many cycles of human-to-human transmission, rather than just a few cycles before the outbreak dies out (as seen in modern Ebola outbreaks).

Database and conclusions

Database. Supplementary Table S1 lists 10 characteristics for each of 25 important 'temperate' (15) and 'tropical' (10) diseases (see Supplementary Note S3 for details of this distinction). Our aim was to select well-defined diseases causing the highest mortality and/or morbidity and hence of the highest historical and evolutionary significance (see Supplementary Note S1 for details of our selection criteria). Of the 25 diseases, we selected 17 because they are the ones assessed by ref. 8 as imposing the heaviest world burdens today

[1]Department of Epidemiology, School of Public Health, University of California, Los Angeles 90095-1772, USA. [2]Division of Infectious Diseases, David Geffen School of Medicine, University of California, Los Angeles 90095-1688, USA. [3]Departments of Geography and of Environmental Health Sciences, University of California, Los Angeles 90095-1524, USA.

Box 1 | Five stages leading to endemic human diseases

We delineate five stages in the transformation of an animal pathogen into a specialized pathogen of humans (Fig. 1). There is no inevitable progression of microbes from Stage 1 to Stage 5: at each stage many microbes remain stuck, and the agents of nearly half of the 25 important diseases we selected for analysis (Supplementary Table S1) have not reached Stage 5.

- **Stage 1.** A microbe that is present in animals but that has not been detected in humans under natural conditions (that is, excluding modern technologies that can inadvertently transfer microbes, such as blood transfusion, organ transplants, or hypodermic needles). Examples: most malarial plasmodia, which tend to be specific to one host species or to a closely related group of host species.
- **Stage 2.** A pathogen of animals that, under natural conditions, has been transmitted from animals to humans ('primary infection') but has not been transmitted between humans ('secondary infection'). Examples: anthrax and tularemia bacilli, and Nipah, rabies and West Nile viruses.
- **Stage 3.** Animal pathogens that can undergo only a few cycles of secondary transmission between humans, so that occasional human outbreaks triggered by a primary infection soon die out. Examples: Ebola, Marburg and monkeypox viruses.
- **Stage 4.** A disease that exists in animals, and that has a natural (sylvatic) cycle of infecting humans by primary transmission from the animal host, but that also undergoes long sequences of secondary transmission between humans without the involvement of animal hosts. We arbitrarily divide Stage 4 into three substages distinguished by the relative importance of primary and secondary transmission:
  Stage 4a. Sylvatic cycle much more important than direct human-to-human spread. Examples: Chagas' disease and (more frequent secondary transmission approaching Stage 4b) yellow fever.
  Stage 4b. Both sylvatic and direct transmission are important. Example: dengue fever in forested areas of West Africa and Southeast Asia.
  Stage 4c. The greatest spread is between humans. Examples: influenza A, cholera, typhus and West African sleeping sickness.
- **Stage 5.** A pathogen exclusive to humans. Examples: the agents causing *falciparum* malaria, measles, mumps, rubella, smallpox and syphilis. In principle, these pathogens could have become confined to humans in either of two ways: an ancestral pathogen already present in the common ancestor of chimpanzees and humans could have co-speciated long ago, when the chimpanzee and human lineages diverged around five million years ago; or else an animal pathogen could have colonized humans more recently and evolved into a specialized human pathogen. Co-speciation accounts well for the distribution of simian foamy viruses of non-human primates, which are lacking and presumably lost in humans: each virus is restricted to one primate species, but related viruses occur in related primate species[19]. While both interpretations are still debated for *falciparum* malaria, the latter interpretation of recent origins is widely preferred for most other human Stage 5 diseases of Supplementary Table S1.

(they have the highest disability-adjusted life years (DALY) scores). Of the 17 diseases, 8 are temperate (hepatitis B, influenza A, measles, pertussis, rotavirus A, syphilis, tetanus and tuberculosis), and 9 are tropical (acquired immune deficiency syndrome (AIDS), Chagas' disease, cholera, dengue haemorrhagic fever, East and West African sleeping sicknesses, *falciparum* and *vivax* malarias, and visceral leishmaniasis). We selected eight others (temperate diphtheria, mumps, plague, rubella, smallpox, typhoid and typhus, plus tropical yellow fever) because they imposed heavy burdens in the past, although modern medicine and public health have either eradicated them (smallpox) or reduced their burden. Except for AIDS, dengue fever, and cholera, which have spread and attained global impact in modern times, most of these 25 diseases have been important for more than two centuries.

Are our conclusions robust to variations in these selection criteria? For about a dozen diseases with the highest modern or historical burdens (for example, AIDS, malaria, plague, smallpox), there can be little doubt that they must be included, but one could debate some of the next choices. Hence we drew up three alternative sets of diseases sharing a first list of 16 indisputable major diseases but differing in the next choices, and we performed all 10 analyses described below on all three sets. It turned out that, with one minor exception, the three sets yielded qualitatively the same conclusions for all 10 analyses, although differing in their levels of statistical significance (see Supplementary Note S4). Thus, our conclusions do seem to be robust.

Temperate/tropical differences. Comparisons of these temperate and tropical diseases yield the following conclusions:

- A higher proportion of the diseases is transmitted by insect vectors in the tropics (8/10) than in the temperate zones (2/15) ($P < 0.005$, χ^2-test, degrees of freedom, d.f. = 1). This difference may be partly related to the seasonal cessations or declines of temperate insect activity.
- A higher proportion ($P = 0.009$) of the diseases conveys long-lasting immunity (11/15) in the temperate zones than in the tropics (2/10).
- Animal reservoirs are more frequent ($P < 0.005$) in the tropics (8/10) than in the temperate zones (3/15). The difference is in the reverse direction ($P = 0.1$, NS, not significant) for environmental reservoirs (1/10 versus 6/15), but those environmental reservoirs that do exist are generally not of major significance except for soil bearing tetanus spores.
- Most of the temperate diseases (12/15) are acute rather than slow, chronic, or latent: the patient either dies or recovers within one to several weeks. Fewer ($P = 0.01$) of the tropical diseases are acute: 3/10 last for one or two weeks, 3/10 last for weeks to months or years, and 4/10 last for many months to decades.
- A somewhat higher proportion of the diseases ($P = 0.08$, NS) belongs to Stage 5 (strictly confined to humans) in the temperate zones (10/15 or 11/15) than in the tropics (3/10). The paucity of Stage 2 and Stage 3 diseases (a total of only 5 such diseases) on our list of 25 major human diseases is noteworthy, because some Stage 2 and Stage 3 pathogens (such as anthrax and Ebola) are notoriously virulent, and because theoretical reasons are often advanced (but also denied) as to why Stage 5 microbes with long histories of adaptation to humans should tend to evolve low morbidity and mortality and not cause major diseases. We discuss explanations for this outcome in Supplementary Note S5.

Most (10/15) of the temperate diseases, but none of the tropical diseases ($P < 0.005$), are so-called 'crowd epidemic diseases' (asterisked in Supplementary Table S1), defined as ones occurring locally as a brief epidemic and capable of persisting regionally only in large human populations. This difference is an immediate consequence of the differences enumerated in the preceding five paragraphs. If a disease is acute, efficiently transmitted, and quickly leaves its victim either dead or else recovering and immune to re-infection, the epidemic soon exhausts the local pool of susceptible potential victims. If in addition the disease is confined to humans and lacks significant animal and environmental reservoirs, depletion of the local pool of potential victims in a small, sparse human population results in local termination of the epidemic. If, however, the human population is large and dense, the disease can persist by spreading to infect people in adjacent areas, and then returning to the original area in a later year, when births and growth have regenerated a new crop of previously unexposed non-immune potential victims. Empirical epidemiological studies of disease persistence or disappearance in isolated human populations of various sizes have yielded estimates of the population required to sustain a crowd disease: at least several hundred thousand people in the cases of measles, rubella and pertussis[2,7]. But human populations of that size did not exist anywhere in the

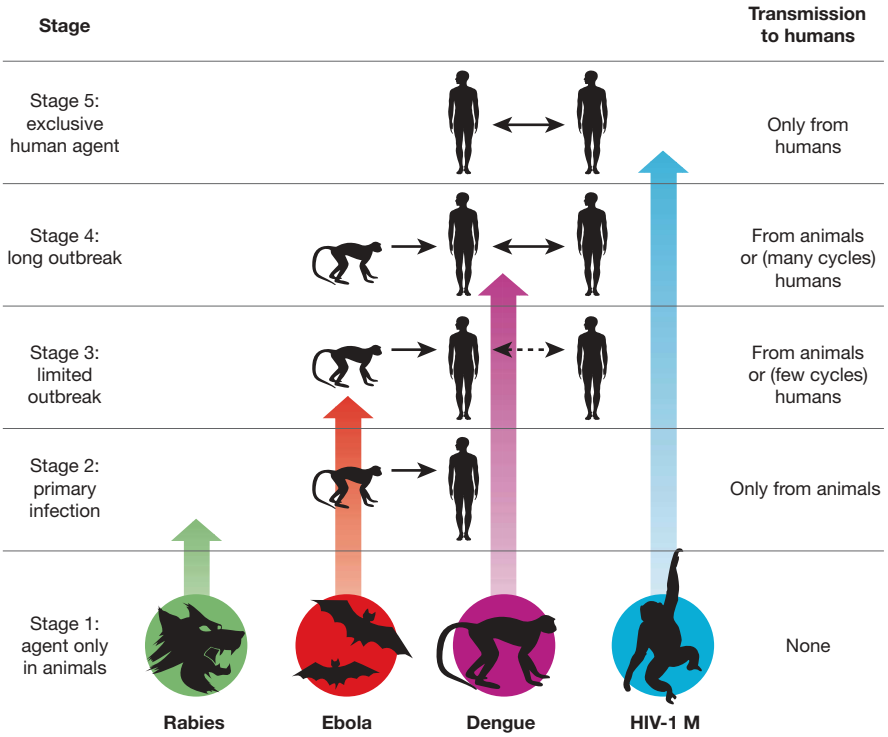


Figure 1 | Illustration of the five stages through which pathogens of animals evolve to cause diseases confined to humans. (See Box 1 for details.) The four agents depicted have reached different stages in the process, ranging from rabies (still acquired only from animals) to HIV-1 (now acquired only from humans).

world until the steep rise in human numbers that began around 11,000 years ago with the development of agriculture[1,9]. Hence the crowd epidemic diseases of the temperate zones must have evolved since then.

Of course, this does not mean that human hunter/gatherer communities lacked infectious diseases. Instead, like the sparse populations of our primate relatives, they suffered from infectious diseases with characteristics permitting them to persist in small populations, unlike crowd epidemic diseases. Those characteristics include: occurrence in animal reservoirs as well as in humans (such as yellow fever); incomplete and/or non-lasting immunity, enabling recovered patients to remain in the pool of potential victims (such as malaria); and a slow or chronic course, enabling individual patients to continue to infect new victims over years, rather than for just a week or two (such as Chagas' disease).

Pathogen origins. (See details for each disease in Supplementary Note S10). Current information suggests that 8 of the 15 temperate diseases probably or possibly reached humans from domestic animals (diphtheria, influenza A, measles, mumps, pertussis, rotavirus, smallpox, tuberculosis); three more probably reached us from apes (hepatitis B) or rodents (plague, typhus); and the other four (rubella, syphilis, tetanus, typhoid) came from still-unknown sources (see Supplementary Note S6). Thus, the rise of agriculture starting 11,000 years ago played multiple roles in the evolution of animal pathogens into human pathogens[1,4,10]. Those roles included both generation of the large human populations necessary for the evolution and persistence of human crowd diseases, and generation of large populations of domestic animals, with which farmers came into much closer and more frequent contact than hunter/gatherers had with wild animals. Moreover, as illustrated by influenza A, these domestic animal herds served as efficient conduits for pathogen transfers from wild animals to humans, and in the process may have evolved specialized crowd diseases of their own.

It is interesting that fewer tropical than temperate pathogens originated from domestic animals: not more than three of the ten tropical diseases of Supplementary Table S1, and possibly none (see Supplementary Note S7). Why do temperate and tropical human diseases differ so markedly in their animal origins? Many (4/10) tropical diseases (AIDS, dengue fever, *vivax* malaria, yellow fever) but only 1/15 temperate diseases (hepatitis B) have wild non-human primate origins ($P = 0.04$). This is because although non-human primates are the animals most closely related to humans and hence pose the weakest species barriers to pathogen transfer, the vast majority of primate species is tropical rather than temperate. Conversely, few tropical but many temperate diseases arose from domestic animals, and this is because domestic animals live mainly in the temperate zones, and their concentration there was formerly even more lop-sided (see Supplementary Note S8).

A final noteworthy point about animal-derived human pathogens is that virtually all arose from pathogens of other warm-blooded vertebrates, primarily mammals plus in two cases (influenza A and ultimately *falciparum* malaria) birds. This comes as no surprise, considering the species barrier to pathogen transfer posed by phylogenetic distance (Box 2). An expression of this barrier is that primates constitute only 0.5% of all vertebrate species but have contributed about 20% of our major human diseases. Expressed in another way, the number of major human diseases contributed, divided by the number of animal species in the taxonomic group contributing those diseases, is approximately 0.2 for apes, 0.017 for non-human primates other than apes, 0.003 for mammals other than primates, 0.00006 for vertebrates other than mammals, and either 0 or else 0.000003 (if cholera really came from aquatic invertebrates) for animals other than vertebrates (see Supplementary Note S9).

Geographic origins. To an overwhelming degree, the 25 major human pathogens analysed here originated in the Old World. That proved to be of great historical importance, because it facilitated the European conquest of the New World (the Americas). Far more Native Americans resisting European colonists died of newly introduced Old World diseases than of sword and bullet wounds. Those invisible agents of New World conquest were Old World microbes to which Europeans had both some acquired immunity based on individual exposure and some genetic resistance based on population

Box 2 | Transitions between stages

Transition from Stage 1 to Stage 2. Most animal pathogens are not transmitted to humans, that is, they do not even pass from Stage 1 to Stage 2. This problem of cross-species infection has been discussed previously[20–23]. Briefly, the probability-per-unit-time (p) of infection of an individual of a new (that is, new recipient) host species increases with the abundance of the existing (that is, existing donor) host, with the fraction of the existing host population infected, with the frequency of 'encounters' (opportunities for transmission, including indirect 'encounters' via vectors) between an individual of the existing host and of the new host, and with the probability of transmission per encounter. p decreases with increasing phylogenetic distance between the existing host and new host. p also varies among microbes (for example, trypanosomes and flaviviruses infect a wide taxonomic range of hosts, while plasmodia and simian foamy viruses infect only a narrow range), and this variation is related to a microbe's characteristics, such as its ability to generate genetic variability, or its ability to overcome host molecular barriers of potential new hosts (such as humoral and cellular defenses or lack of cell membrane receptors essential for microbe entry into host cells).

These considerations illuminate different reasons why a given animal host species may or may not become a source of many infections in humans. For instance, despite chimpanzees' very low abundance and infrequent encounters with humans, they have donated to us numerous zoonoses (diseases that still mainly afflict animals) and one or two established human diseases (AIDS and possibly hepatitis B) because of their close phylogenetic relationship to humans. Despite their large phylogenetic distance from humans, many of our zoonoses and probably two of our established diseases (plague and typhus) have been acquired from rodents, because of their high abundance and frequent encounters with humans in dwellings. Similarly, about half of our established temperate diseases have been acquired from domestic livestock, because of high local abundance and very frequent contact. Conversely, elephants and bats are not known to have donated directly to us any established diseases and rarely donate zoonoses, because they are heavily penalized on two or three counts: large phylogenetic distance, infrequent encounters with humans, and (in the case of elephants) low abundance. One might object that Nipah, severe acute respiratory syndrome (SARS) and rabies viruses do infect humans from bats, but these apparent exceptions actually support our conclusion. While bats may indeed be the primary reservoir for Nipah and SARS, human infections by these viruses are acquired mainly from intermediate animal hosts that frequently encounter humans (respectively, domestic pigs, and wild animals sold for food). The rare cases of rabies transmission directly to humans from bats arise because rabies changes a bat's behaviour so that it does encounter and bite humans, which a healthy bat (other than a vampire bat) would never do.

Transition from Stage 2 to Stage 3 or 4. Although some Stage 2 and 3 pathogens, such as the anthrax and Marburg agents, are virulent and feared, they claim few victims at present. Yet if they made the transition to Stage 4 or 5, their global impact would be devastating. Why do animal pathogens that have survived the initial jump across species lines into a human host (Stages 1 to 2) usually reach a dead end there, and not evolve past Stages 3 and 4 into major diseases confined to humans (Stage 5)? Barriers between Stages 2 and 3 (consider the rabies virus) include differences between human and animal behaviour affecting transmission (for example, animals often bite humans but humans rarely bite other humans); a pathogen's need to evolve adaptations to the new human host and possibly also to a new vector; and obstacles to a pathogen's spread between human tissues (for example, BSE is restricted to the central nervous system and lymphoid tissue). Barriers between Stages 3 and 4 (consider Ebola virus) include those related to human population size and to transmission efficiency between humans. The emergence of novel pathogens is now being facilitated by modern developments exposing more potential human victims and/or making transmission between humans more efficient than before[24–27]. These developments include blood transfusion (hepatitis C), the commercial bushmeat trade (retroviruses), industrial food production (bovine spongiform encephalitis, BSE), international travel (cholera), intravenous drug use (HIV), vaccine production (simian virus 40, SV40), and susceptible pools of elderly, antibiotic-treated, immunosuppressed patients (see Supplementary Note S2 for details).

exposure over time, but to which previously unexposed Native American populations had no immunity or resistance[1,4–6]. In contrast, no comparably devastating diseases awaited Europeans in the New World, which proved to be a relatively healthy environment for Europeans until yellow fever and malaria of Old World origins arrived[11].

Why was pathogen exchange between Old and New Worlds so unequal? Of the 25 major human diseases analysed, Chagas' disease is the only one that clearly originated in the New World. For two others, syphilis and tuberculosis, the debate is unresolved: it remains uncertain in which hemisphere syphilis originated, and whether tuberculosis originated independently in both hemispheres or was brought to the Americas by Europeans. Nothing is known about the geographic origins of rotavirus, rubella, tetanus and typhus. For all of the other 18 major pathogens, Old World origins are certain or probable.

Our preceding discussion of the animal origins of human pathogens may help explain this asymmetry. More temperate diseases arose in the Old World than New World because far more animals that could furnish ancestral pathogens were domesticated in the Old World. Of the world's 14 major species of domestic mammalian livestock, 13, including the five most abundant species with which we come into closest contact (cow, sheep, goat, pig and horse), originated in the Old World[1]. The sole livestock species domesticated in the New World was the llama, but it is not known to have infected us with any pathogens[1,2]—perhaps because its traditional geographic range was confined to the Andes, it was not milked or ridden or hitched to ploughs, and it was not cuddled or kept indoors (as are some calves, lambs and piglets). Among the reasons why far more tropical diseases (nine versus one) arose in the Old World than the New World are that the genetic distance between humans and New World monkeys is almost double that between humans and Old World monkeys, and is many times that between humans and Old World apes; and that much more evolutionary time was available for transfers from animals to humans in the Old World (about 5 million years) than in the New World (about 14,000 years).

Outlook and future research directions

Many research directions on infectious disease origins merit more effort. We conclude by calling attention to two such directions: clarifying the origins of existing major diseases, and surveillance for early detection of new potentially major diseases.

Origins of established diseases. This review illustrates big gaps in our understanding of the origins of even the established major infectious diseases. Almost all the studies that we have reviewed were based on specimens collected opportunistically from domestic animals and a few easily sampled wild animal species, rather than on systematic surveys for particular classes of agents over the spectrum of domestic and wild animals. A case in point is our ignorance even about smallpox virus, the virus that has had perhaps the greatest impact on human history in the past 4,000 years. Despite some knowledge of poxviruses infecting our domestic mammals, we know little about poxvirus diversity among African rodents, from which those poxviruses of domestic mammals are thought to have evolved. We do not even know whether 'camelpox', the closest known relative of smallpox virus, is truly confined to camels as its name implies or is instead a rodent virus with a broad host range. There could be still-unknown poxviruses more similar to smallpox virus in yet unstudied animal reservoirs, and those unknown poxviruses could be important not only as disease threats but also as reagents for drug and vaccine development.

Equally basic questions arise for other major pathogens. While *falciparum* malaria, an infection imposing one of the heaviest global burdens today, seems to have originated from a bird parasite whose descendants include both the *Plasmodium falciparum* infecting humans and the *P. reichenowii* infecting chimpanzees, malaria researchers still debate whether the bird parasite was introduced to

both humans and chimpanzees[12] a few thousand years ago in association with human agriculture, or instead more than five million years ago before the split of humans and chimpanzees from each other[13]. Although resolving this debate will not help us eradicate malaria, it is fascinating in its own right and could contribute to our broader understanding of disease emergence. In the case of rubella, a human crowd disease that must have emerged only in the past 11,000 years and for which some close relative may thus still exist among animals, no even remotely related virus is known; one or more may be lurking undiscovered somewhere. Does the recent identification of porcine rubulavirus and the Mapuera virus in bats as the closest known relatives of mumps virus mean that pigs infected humans, or that human mumps infected pigs, or that bats independently infected both humans and pigs? Is human tuberculosis descended from a ruminant mycobacterium that recently infected humans from domestic animals (a formerly prevalent view), or from an ancient human mycobacterium that has come to infect domestic and wild ruminants (a currently popular view)?

To fill these and other yawning gaps in our understanding of disease origins, we propose an 'origins initiative' aimed at identifying the origins of a dozen of the most important human infectious diseases: for example, AIDS, cholera, dengue fever, *falciparum* malaria, hepatitis B, influenza A, measles, plague, rotavirus, smallpox, tuberculosis and typhoid. Although more is already known about the origins of some of these agents (AIDS, influenza A and measles) than about others (rotavirus, smallpox and tuberculosis), more comprehensive screening is still likely to yield significant new information about even the most studied agents, as illustrated by the recent demonstration that gorillas rather than chimpanzees were probably the donor species for the O-group of human immunodeficiency virus (HIV)-1[14]. The proposed effort would involve systematic sampling and phylogeographic analysis of related pathogens in diverse animal species: not just pigs and other species chosen for their ready availability, but a wider range of wild and domestic species whose direct contact (for example, as bushmeat) or indirect contact (for example, vector-mediated) with humans could plausibly have led to human infections. In addition to the historical and evolutionary significance of knowledge gained through such an origins initiative, it could yield other benefits such as: identifying the closest relatives of human pathogens; a better understanding of how diseases have emerged; new laboratory models for studying public health threats; and perhaps clues that could aid in predictions of future disease threats.

A global early warning system. Most major human infectious diseases have animal origins, and we continue to be bombarded by novel animal pathogens. Yet there is no ongoing systematic global effort to monitor for pathogens emerging from animals to humans. Such an effort could help us to describe the diversity of microbial agents to which our species is exposed; to characterize animal pathogens that might threaten us in the future; and perhaps to detect and control a local human emergence before it has a chance to spread globally.

In our view, monitoring should focus on people with high levels of exposure to wild animals, such as hunters, butchers of wild game, wildlife veterinarians, workers in the wildlife trade, and zoo workers. Such people regularly become infected with animal viruses, and their infections can be monitored over time and traced to other people in contact with them. One of us (N.D.W.) has been working in Cameroon to monitor microbes in people who hunt wild game, in other people in their community, and in their animal prey[15]. The study is now expanding to other continents and to monitor domestic animals (such as dogs) that live in close proximity to humans but are exposed to wild animals through hunting and scavenging. Monitoring of people, animals, and animal die-offs[16] will serve as an early warning system for disease emergence, while also providing a unique archive of pathogens infecting humans and the animals to which we are exposed. Specimens from such highly exposed human populations could be screened specifically for agents known to be present in the animals they hunt (for example, retroviruses among hunters of non-human primates), as well as generically using broad screening tools such as viral microarrays[17] and random amplification polymerase chain reaction (PCR)[18]. Such monitoring efforts also provide potentially invaluable repositories, which would be available for study after future outbreaks in order to reconstruct an outbreak's origin, and as a source of relevant reagents.

1. Diamond, J. *Guns, Germs, and Steel: the Fates of Human Societies* (Norton, New York, 1997).
2. Dobson, A. P. & Carper, E. R. Infectious diseases and human population history. *Bioscience* **46,** 115–126 (1996).
3. Diamond, J. & Panosian, C. in *When Disease Makes History: Epidemics and Great Historical Turning Points* (ed. Hämäläinen, P.) 17–44 (Helsinki Univ. Press, 2006).
4. McNeill, W. H. *Plagues and Peoples* (Anchor, Garden City, 1976).
5. Crosby, A. W. *Ecological Imperialism: the Biological Expansion of Europe* 900–1900 (Cambridge Univ. Press, Cambridge, UK, 1986).
6. Ramenofsky, A. *Vectors of Death: the Archaeology of European Contact* (New Mexico Press, Albuquerque, 1987).
7. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, Oxford, UK, 1991).
8. Lopez, A. D., Mathers, C. D., Ezzati, N., Jamison, D. T. & Murray, C. J. L. (eds) *Global Burden of Disease and Risk Factors* (Oxford Univ. Press, New York, 2006).
9. Bellwood, P. *First Farmers: the Origins of Agriculture Societies* (Blackwell, Oxford, 2005).
10. Diamond, J. Evolution, consequences, and future of plant and animal domestication. *Nature* **418,** 34–41 (2002).
11. McNeill, J. R. in *When Disease Makes History: Epidemics and Great Historical Turning Points* (ed. Hämäläinen, P.) 81–111 (Helsinki Univ. Press, Helsinki, 2006).
12. Waters, A. P., Higgins, D. G. & McCutchan, T. F. *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. *Proc. Natl Acad. Sci. USA* **88,** 3140–3144 (1991).
13. Ayala, F. J., Escalante, A. A. & Rich, S. M. Evolution of *Plasmodium* and the recent origin of the world populations of *Plasmodium falciparum*. *Parassitologia* **41,** 55–68 (1999).
14. Van Heuverswyn, F. *et al.* Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature* **444,** 164 (2006).
15. Wolfe, N. D. *et al.* Naturally acquired simian retrovirus infections in central African hunters. *Lancet* **363,** 932–937 (2004).
16. Kuiken, T. *et al.* Pathogen surveillance in animals. *Science* **309,** 1680–1681 (2005).
17. Wang, D. *et al.* Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol.* **1,** E2 (2003).
18. Jones, M. S. *et al.* New DNA viruses identified in patients with acute viral infection syndrome. *J. Virol.* **79,** 8230–8236 (2005).
19. Switzer, W. M. *et al.* Ancient co-speciation of simian foamy viruses and primates. *Nature* **434,** 376–380 (2005).
20. Taylor, L. H., Latham, S. M. & Woolhouse, M. E. Risk factors for human disease emergence. *Phil. Trans. R. Soc. Lond. B* **356,** 983–989 (2001).
21. Moya, A., Holmes, E. C. & Gonzalez-Candelas, F. The population genetics and evolutionary epidemiology of RNA viruses. *Nature Rev. Microbiol.* **2,** 279–288 (2004).
22. Antia, R., Regoes, R. H., Koella, J. C. & Bergstrom, C. T. The role of evolution in the emergence of infectious diseases. *Nature* **426,** 658–661 (2003).
23. May, R. M., Gupta, S. & McLean, A. R. Infectious disease dynamics: what characterizes a successful invader? *Phil. Trans. R. Soc. Lond. B* **356,** 901–910 (2001).
24. Morens, D. M., Folkers, G. K. & Fauci, A. S. The challenge of emerging and re-emerging infectious diseases. *Nature* **430,** 242–249 (2004).
25. Morse, S. S. Factors in the emergence of infectious diseases. *Emerg. Infect. Dis.* **1,** 7–15 (1995).
26. Wilson, M. E. Travel and the emergence of infectious diseases. *Emerg. Infect. Dis.* **1,** 39–46 (1995).
27. Weiss, R. A. & McMichael, A. J. Social and environmental risk factors in the emergence of infectious diseases. *Nature Med.* **10,** S70–S76 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Krain for assistance with Supplementary Note S10; M. Antolin, D. Burke, L. Fleisher, E. Holmes, L. Real, A. Rimoin, R. Weiss and M. Woolhouse for comments; and many other colleagues for providing information. This work was supported by an NIH Director's Pioneer Award and Fogarty International Center IRSDA Award (to N.D.W.), a W. W. Smith Foundation award (to N.D.W.), and National Geographic Society awards (to J.D. and N.D.W.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to N.W. (nwolfe@ucla.edu) or J.D. (jdiamond@geog.ucla.edu).

ARTICLES

A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity

Zheng Qing Fu[1]*, Ming Guo[1]*, Byeong-ryool Jeong[1], Fang Tian[1,2], Thomas E. Elthon[2,3], Ronald L. Cerny[4], Dorothee Staiger[5] & James R. Alfano[1]

The bacterial plant pathogen *Pseudomonas syringae* injects effector proteins into host cells through a type III protein secretion system to cause disease. The enzymatic activities of most of *P. syringae* effectors and their targets remain obscure. Here we show that the type III effector HopU1 is a mono-ADP-ribosyltransferase (ADP-RT). HopU1 suppresses plant innate immunity in a manner dependent on its ADP-RT active site. The HopU1 substrates in *Arabidopsis thaliana* extracts were RNA-binding proteins that possess RNA-recognition motifs (RRMs). *A. thaliana* knockout lines defective in the glycine-rich RNA-binding protein GRP7 (also known as AtGRP7), a HopU1 substrate, were more susceptible than wild-type plants to *P. syringae*. The ADP-ribosylation of GRP7 by HopU1 required two arginines within the RRM, indicating that this modification may interfere with GRP7's ability to bind RNA. Our results suggest a pathogenic strategy where the ADP-ribosylation of RNA-binding proteins quells host immunity by affecting RNA metabolism and the plant defence transcriptome.

Many Gram-negative pathogens of plants and animals and other eukaryotic-associated bacteria use type III protein secretion systems[1,2]. Type III protein secretion systems are molecular syringes that inject bacterial proteins called effectors into eukaryotic host cells to modulate host physiology. In animal cells, their activities alter specific host cell functions, including phagocytosis, proinflammatory responses, apoptosis and intracellular trafficking[3]. Much less is understood about the activities and targets of type III effectors from plant pathogens. The emerging picture is that many type III effectors from plant pathogens suppress host immune responses[4–6]. Thus far, effectors that possess cysteine protease, tyrosine phosphatase and E3 ubiquitin ligase activities have been implicated in suppression of plant innate immunity[7–13]; however, the enzymatic activities for most plant pathogen type III effectors that suppress innate immunity remain unknown.

Genomic investigations of *Pseudomonas syringae* pv. tomato DC3000, a pathogen of *Arabidopsis thaliana* and tomato, have identified greater than 30 effector genes[14,15]. Among these, *hopO1-1*, *hopU1* and *hopO1-2* (formerly *hopPtoS1*, *hopPtoS2*, and *hopPtoS3*, respectively) are predicted to encode proteins that contain potential active sites of mono-ADP ribosyltransferases (ADP-RTs)[16,17] (Fig. 1a). ADP-RTs are well-characterized toxins in animal pathogens, including two that are type-III-injected[18,19], but they have not been demonstrated to be important in plant pathogenicity. Furthermore, genes that encode ADP-RTs have been found in eukaryotes, but have not been identified in plants[20].

hopU1 encodes a putative ADP-RT

To begin to characterize the DC3000 effector genes that potentially encode ADP-RTs we focused on *hopU1*, which encodes a product with similarity to known ADP-RTs (Fig. 1a). *hopU1* is downstream of an apparent type III promoter, of the *shcF* type III chaperone gene and of the *hopF2* effector gene in the DC3000 chromosome (Fig. 1b). Semi-quantitative reverse transcriptase polymerase chain reaction (RT–PCR) experiments indicated the *hopU1* gene is transcribed and its expression was elevated when DC3000 was grown in a medium that induces the expression of the type III protein secretion system (Fig. 1c) and HopU1 was shown to be type III-injected into plant cells on the basis of adenylate cyclase translocation assays (Supplementary Fig. 1a). A DC3000 Δ*hopU1* mutant was reduced sixfold in its ability to multiply in plant tissue and cause disease symptoms in *A. thaliana* ecotype Col-0 (Supplementary Fig. 1b).

HopU1 suppresses plant innate immunity

We earlier reported that DC3000 mutants defective in type III effectors that can suppress the hypersensitive response—a programmed cell death of plant cells associated with innate immunity—often show an enhanced ability to elicit a hypersensitive response[21]. To investigate whether the Δ*hopU1* mutant shared this phenotype we infiltrated wild-type DC3000 and the Δ*hopU1* mutant at different cell densities into *Nicotiana tabacum* cv. Xanthi (tobacco).We consistently found that the Δ*hopU1* mutant elicited a hypersensitive response in tobacco at cell densities below the threshold needed for wild-type DC3000 (Fig. 2a). When *hopU1* was expressed *in trans* in the Δ*hopU1* mutant it complemented this phenotype (Fig. 2a). We also assessed cell death by measuring the amount of ion leakage from plant cells and found that the Δ*hopU1* mutant caused more cell death than DC3000 and expression of *hopU1 in trans* reduced the amount of cell death to wild-type levels (Supplementary Fig. 2a). Taken together, these results provide genetic evidence that HopU1 acts as a suppressor of the non-host hypersensitive response.

[1]Plant Science Initiative and Department of Plant Pathology, University of Nebraska, Lincoln, Nebraska 68588-0660, USA. [2]School of Biological Sciences, University of Nebraska, Lincoln, Nebraska 68588-0118, USA. [3]Department of Agronomy and Horticulture, University of Nebraska, Lincoln, Nebraska 68588-0915, USA. [4]Department of Chemistry, University of Nebraska, Lincoln, Nebraska 68588-0304, USA. [5]Molecular Cell Physiology, University of Bielefeld, 33501 Bielefeld, Germany.
*These authors contributed equally to this work.

To determine whether the predicted ADP-RT activity of HopU1 was required for suppression of the hypersensitive response, we ectopically expressed in the Δ*hopU1* mutant a HopU1 derivative ($HopU1_{DD}$) that had glutamic acid residues substituted with aspartic acids in the putative ADP-RT active site (Fig. 1a). This strain elicited an enhanced hypersensitive response similar to the Δ*hopU1* mutant control, suggesting that the suppression of the hypersensitive response required a functional ADP-RT active site (Fig. 2b). Ion leakage conductivity assays indicated that the Δ*hopU1* mutant expressing $HopU1_{DD}$ caused similar amounts of cell death to the Δ*hopU1* mutant control (Supplementary Fig. 2b). Thus, the hypersensitive response suppression activity of HopU1 requires a functional ADP-RT catalytic site.

We reasoned that HopU1 may be capable of suppressing other innate immune responses. To test this, we generated transgenic *A. thaliana* Col-0 plants that constitutively produced HopU1 fused to a haemagglutinin epitope (HopU1–HA). Bacterial flagellin often acts as a pathogen-associated molecular pattern, a conserved molecule from a microorganism recognized by animal and plant innate immune systems[22–24]. A conserved peptide from bacterial flagellin, flg22, has been shown to be effective at triggering callose (β-1-3 glucan) deposition[25]. The HopU1–HA-expressing transgenic plants treated with flg22 produced significantly reduced amounts of callose compared with wild-type plants (Fig. 2c, d). These plants also elicited a delayed atypical hypersensitive response in response to the type III effector AvrRpt2, which is recognized by the RPS2 resistance protein present in *A. thaliana* Col-0 (ref. 26; Supplementary Fig. 3). Together, these data indicate that HopU1 can suppress innate immune responses that are triggered by either a type III effector protein or a pathogen-associated molecular pattern.

HopU1 is an active ADP-RT

To explore whether HopU1 indeed possessed ADP-RT activity, we purified recombinant HopU1 and the catalytic site mutant ($HopU1_{DD}$), both fused to histidine affinity tags (Fig. 3a). The activity of recombinant HopU1–His was tested with poly-L-arginine, an artificial substrate for many ADP-RTs that can modify arginine residues. HopU1–His was capable of ADP-ribosylating poly-arginine in the presence of [^{32}P]-NAD, whereas the $HopU1_{DD}$–His mutant incorporated radioactivity in amounts similar to the BSA control (Fig. 3b). Therefore, HopU1 is an active ADP-RT that can modify arginine residues.

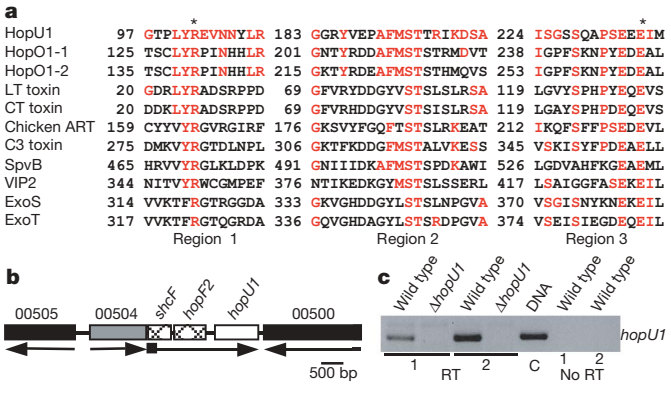


Figure 1 | HopU1 is a putative mono-ADP-ribosyltransferase that contributes to virulence. **a**, Alignment of the conserved regions of known mono-ADP-ribosyltransferases (ADP-RTs) with putative DC3000 ADP-RTs. Conserved residues are shown in red and the invariant amino acids of the cholera toxin group of ADP-RTs[42] are marked with asterisks. **b**, *hopU1* (white box) is downstream of a type-III-related promoter, the *shcF* type III chaperone gene and the *hopF2* effector gene (hatched boxes). **c**, RNA was isolated from DC3000 (wild type; WT) or the Δ*hopU1* mutant grown in either rich media (1) or a minimal medium that induces type-III-related genes (2) and used in RT–PCR reactions. A DNA control (C) and no reverse transcriptase controls (No RT) were included.

We next examined whether HopU1–His was capable of using plant proteins as substrates. Crude protein extracts from the leaves of *A. thaliana* ecotype Col-0 and tobacco were used in ADP-RT reactions. ADP-RT reactions were separated by SDS–polyacrylamide gel electrophoresis (PAGE), and subjected to autoradiography (Fig. 3c). At least two proteins in *A. thaliana* extracts and three in tobacco were ADP-ribosylated by HopU1–His. No labelled products were detected from reactions using the inactive $HopU1_{DD}$–His (Fig. 3c). Therefore, HopU1–His can use several *Arabidopsis* and tobacco proteins as substrates.

RNA-binding proteins are HopU1 targets

To determine the identity of the proteins ADP-ribosylated by HopU1 in *A. thaliana* extracts, we separated ADP-RT reactions with two-dimensional (2D) PAGE followed by autoradiography (for representative examples, see Supplementary Fig. 4). One well-separated protein spot that stained with Coomassie blue had identical migration to a radiolabelled ADP-RT activity spot. It was analysed with tandem mass spectrometry and corresponded to chloroplast RNA-binding protein (CP-RBP) RBP31 (ref. 27; Supplementary Table 1). The other ADP-RT activity spots did not co-migrate with visible Coomassie-blue-stained protein spots, suggesting that these proteins

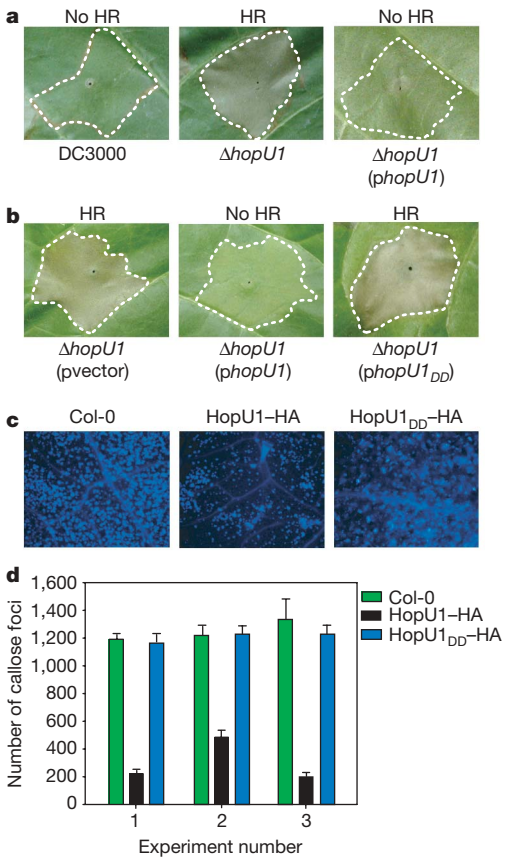


Figure 2 | HopU1 suppresses outputs of plant innate immunity. **a**, DC3000, Δ*hopU1* mutant and the Δ*hopU1* mutant expressing *hopU1* (p*hopU1*) were infiltrated into tobacco leaves at threshold cell densities (1×10^6 cells per ml). After 24 h the tissue was assessed for hypersensitive response (HR) production. **b**, The Δ*hopU1* mutant carrying a vector control (pvector), a *hopU1* construct (p*hopU1*) or a *hopU1* ADP-RT catalytic mutant construct (p$hopU1_{DD}$) were infiltrated into tobacco at 1×10^7 cells per ml and assessed for hypersensitive response production after 24 h. **c**, Callose deposition was visualized in *A. thaliana* plants expressing HopU1–HA or the $HopU1_{DD}$–HA mutant 16 h after treatment with flg22. **d**, Callose deposition was quantified by counting the number of callose foci per field of view. Twenty leaf regions (4 fields of view from 5 different leaves) were averaged and error bars (s.e.m.) are indicated. The experiment was repeated three times.

were in lower abundance. To enrich for less abundant substrates more concentrated *A. thaliana* extracts were made and fractionated using ion exchange chromatography. ADP-RT assays were performed on aliquots of each fraction (Supplementary Fig. 4b). Coomassie-blue-stained protein spots that had identical migration patterns to ADP-RT-modified protein spots visible on autoradiograms were analysed with tandem mass spectrometry. In total, we found three CP-RBPs and two glycine-rich RNA-binding proteins (GR-RBPs), GRP7 and GRP8 (ref. 28), to have been ADP-ribosylated by HopU1–His (Supplementary Table 1).

To confirm that the identified CP-RBPs and GR-RBPs were substrates for HopU1–His, we constructed glutathione *S*-transferase (GST) fusions. Substrate–GST fusions were partially purified with glutathione Sepharose or used as crude lysates in ADP-RT reactions. Each of the substrate–GST fusions could act as substrates for HopU1–His (Fig. 4a). Other plant protein–GST fusions corresponding to proteins neighbouring HopU1 substrates on 2D polyacrylamide gels were not ADP-ribosylated by HopU1–His. Thus, these CP-RBPs and GR-RBPs are indeed *in vitro* substrates for the HopU1 ADP-RT. We did not find additional HopU1 substrates in our extensive mass spectrometry analyses, suggesting that the *in vivo* targets of HopU1 are among these proteins.

CP-RBPs and GR-RBPs belong to a group of RNA-binding proteins that all have in common an RNA-recognition motif (RRM)—a protein domain demonstrated to be necessary and sufficient for RNA binding[29,30]. Localization experiments found that HopU1–green fluorescent protein (GFP) (and HopU1–GUS) and the GR-RBP GRP7–GFP and GRP8–GFP fusion proteins were similarly localized to the cytoplasm and possibly to the nucleus, whereas the CP-RBP–GFP fusions were discretely localized to the chloroplast (Supplementary Fig. 5). Because of our localization experiments and the association of GR-RBPs with abiotic and biotic stress[31,32], we focused on the GR-RBPs as putative physiological targets of HopU1.

The GR-RBP GRP7 and GRP8 are homologous to each other, sharing 76.9% identity, and probably perform related functions. GRP7 has been shown to bind RNA and influence messenger RNA oscillations in response to circadian rhythms at the post-transcriptional level[33,34]. GRP7 contains 14 arginine residues that represent putative sites of ADP-ribosylation (Fig. 4b). We found that the RRM–GST fusion was ADP-ribosylated by HopU1–His, whereas the glycine-rich-domain–GST fusion was not (Fig. 4c). To determine the arginine residues required for ADP-ribosylation, each arginine of the RRM was individually mutated to lysine in full-length GRP7–GST fusions. When arginines in positions 47 or 49 of GRP7 were substituted with

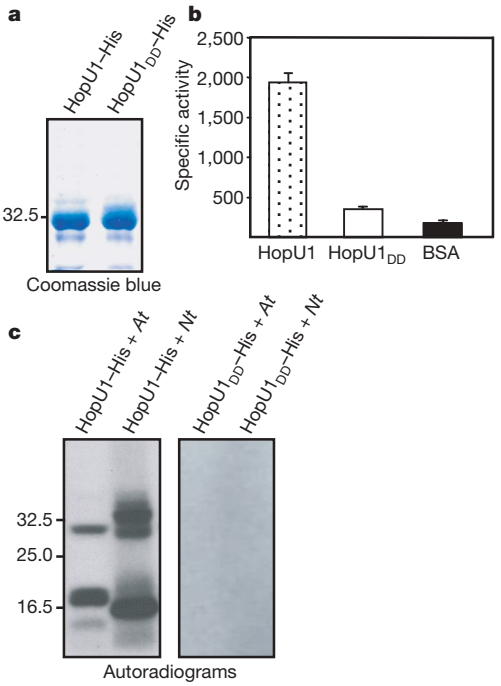


Figure 3 | HopU1–His ADP-ribosylates poly-L-arginine and proteins in *Arabidopsis* and tobacco. **a**, SDS–PAGE of partially purified HopU1–His and HopU1–His catalytic mutant ($HopU1_{DD}$–His) used in ADP-RT assays. **b**, HopU1–His (stippled bar), $HopU1_{DD}$–His (white bar), and a BSA control (black bar) were incubated with poly-L-arginine in the presence of [^{32}P]-NAD. ^{32}P-labelled products were quantified with liquid scintillation. The specific activity unit is μmol ^{32}P transferred per min $\times 10^{-10}$ per μmol of HopU1. The experiment was performed twice and the standard errors are indicated. **c**, Autoradiograms of ADP-RT assays with either *Arabidopsis* (*At*) or tobacco (*Nt*) extracts with HopU1–His or $HopU1_{DD}$–His. **a**, **c**, Molecular mass markers (kDa) are indicated.

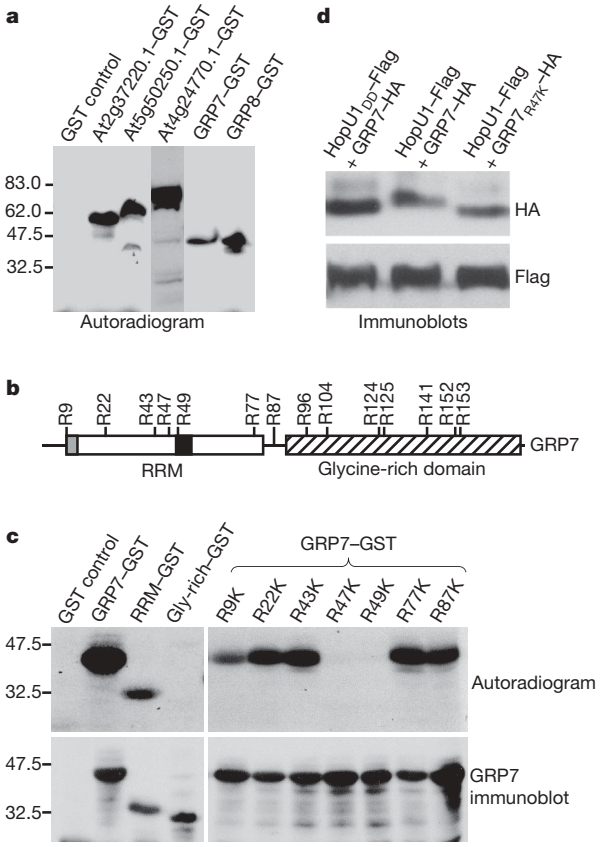


Figure 4 | HopU1–His ADP-ribosylates recombinant RNA-binding proteins *in vitro* and *in planta*. **a**, Recombinant CP-RBP and GR-RBP glutathione *S*-transferase (GST) fusions and a GST control were used in ADP-RT assays with HopU1–His. ADP-RT reactions were subjected to SDS–PAGE followed by autoradiography. The substrate fusions are listed with their corresponding *A. thaliana* locus or the protein name. **b**, Schematic representation of the RRM and glycine-rich domains within GRP7 (also known as AtGRP7) and the locations of the arginine residues (R) throughout the protein. Black and grey sections within the RRM domain depict RNP1 and RNP2, respectively, two conserved regions within RRM domains[29]. The locations of the arginine residues (R) are indicated. **c**, *In vitro* ADP-RT assays were done with HopU1–His and various GRP7–GST fusions. These included an RRM–GST fusion, a glycine-rich-domain–GST fusion, and GRP7–GST mutants that have each arginine within the RRM domain separately substituted with lysine. Autoradiograms (top panel) indicate the GRP7–GST derivatives that maintained the ability to be ADP-ribosylated and immunoblots (bottom panel) indicate the relative stability of each GRP7–GST derivative. **d**, GRP7 fused to an HA epitope (GRP7–HA), or a GRP7–HA derivative that is unable to be ADP-ribosylated ($GRP7_{R47K}$–HA) were transiently expressed in *N. benthamiana* with HopU1–Flag or the ADP-RT catalytic mutant $HopU1_{DD}$–Flag. After 40 h, leaf samples were subjected to immunoblot analysis using anti-HA and anti-Flag antibodies. **a**, **c**, Molecular mass markers (kDa) are indicated.

lysine, these GRP7–GST derivatives were no longer ADP-ribosylated, suggesting that one of these residues is the site of the ADP-ribose modification, whereas the other may be required for substrate recognition (Fig. 4c). Interestingly, on the basis of the RRM domain structure in other RNA-binding proteins, both of these residues would probably be solvent-exposed[35]. Moreover, the arginine in position 49 is within RNP1, the most conserved region of the RRM domain and one that has been directly implicated in RNA-binding[30].

To determine if GRP7 can be ADP-ribosylated by HopU1 *in planta*, we co-expressed HopU1–Flag or $HopU1_{DD}$–Flag and GRP7–HA or $GRP7_{R47K}$–HA (a GRP7 derivative that cannot be ADP-ribosylated) in *N. benthamiana* using *Agrobacterium* transient assays. After 40 h, plant extracts isolated from leaf tissue were separated on SDS–PAGE gels and analysed with immunoblots by using anti-Flag or anti-HA antibodies. We consistently observed an increase in the molecular mass of GRP7–HA when it was expressed *in planta* with HopU1–Flag, but not when expressed with $HopU1_{DD}$–Flag or when HopU1–Flag was expressed with $GRP7_{R47K}$–HA (Fig. 4d), suggesting that the increased molecular mass was due to ADP-ribosylation. Therefore, HopU1–Flag can ADP-ribosylate GRP7–HA inside the plant cell.

A *grp7* mutant is more susceptible to *P. syringae*

To determine the involvement of GRP7 in plant innate immunity, we identified an *A. thaliana* SALK homozygous T-DNA insertion line[36] in the *GRP7* locus, and confirmed that it did not produce *GRP7* mRNA and protein (Supplementary Fig. 6). To determine if this mutant, designated *grp7-1*, was altered in its responses to *P. syringae*, we infected wild-type *A. thaliana* Col-0 and the *grp7-1* mutant plants with DC3000 and a Δ*hrcC* mutant defective in the type III protein secretion system. DC3000 caused enhanced disease symptoms on *grp7-1* mutant plants compared with wild-type Col-0 (Fig. 5a). Each strain grew to higher levels in *grp7-1* plants compared with wild-type Col-0 plants (Fig. 5b), indicating that *grp7-1* plants were more susceptible to *P. syringae*. The growth difference was even more pronounced for the Δ*hrcC* mutant, which is probably due to the fact that this strain cannot inject any type III effectors, many of which suppress innate immunity. Importantly, we found that flg22-induced callose deposition was reduced in *grp7-1* plants compared with wild-type *A. thaliana* Col-0 (Fig. 5c), further supporting the idea that *grp7-1* mutant plants were impaired in their innate immune responses. Similar phenotypes were observed for an independent T-DNA mutant designated *grp7-2* (Supplementary Fig. 7).

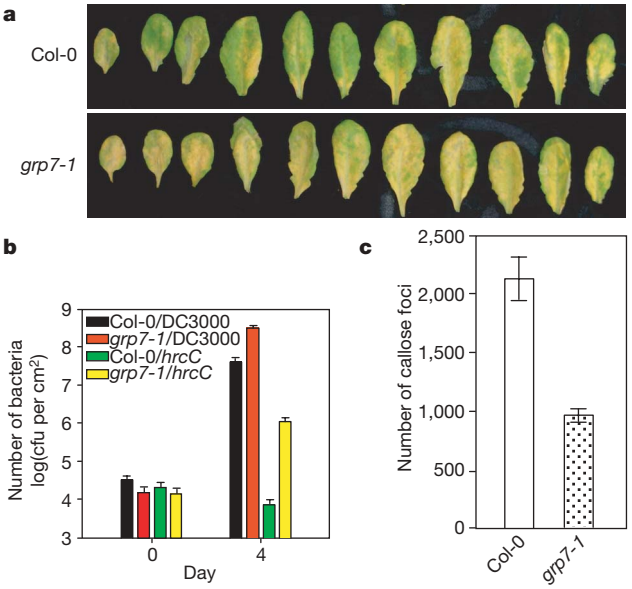


Figure 5 | Analyses of *A. thaliana grp7-1* mutant plants suggest GRP7 has a role in innate immunity. **a**, Disease symptoms in *grp7-1* plants and *A. thaliana* Col-0 plants after spray-inoculation with a cell density of 2×10^8 cells per ml. Pictures were taken after 5 days. The experiments in **a** and **b** were repeated five times with similar results. **b**, Bacterial growth assays of wild-type DC3000 and the Δ*hrcC* mutant spray-inoculated at a cell density of 2×10^8 cells per ml onto *grp7-1* and wild-type *A. thaliana* Col-0 plants. Results are shown as mean ± s.e.m. cfu, colony forming units. **c**, Callose deposition was determined in *A. thaliana* Col-0 and *grp7-1* mutant plants 16 h after treatment with flg22. The number of callose foci per field of view for 20 leaf regions (4 fields of view from 5 different leaves) were averaged ± s.e.m. The experiment was repeated three times with similar results.

Discussion

Several plant pathogen type III effectors are known to suppress outputs of innate immunity[4]; however, the enzymatic activity and substrates for these type III effectors remain poorly understood[14]. Indeed, there are only two cases for which both an enzymatic activity and a substrate for a plant pathogen type III effector are known[7–9]. The substrates of the HopU1 ADP-RT suggest an additional strategy used by bacterial pathogens to modulate plant innate immunity by indirectly affecting host RNA status. That is, GR-RBPs may act as key post-transcriptional regulators through either the trafficking, stabilization or processing of specific mRNAs in response to pathogen stress, and the ADP-ribosylation of the GR-RBPs by HopU1 may disrupt their activity (Supplementary Fig. 8). By disabling the function of GR-RBPs the pathogen may reduce the amount of immunity-related mRNAs available in the plant and tip the balance of the interaction in favour of the pathogen.

To our knowledge, this is the first demonstration of an active ADP-RT virulence protein in a plant pathogen. The DC3000 genome encodes at least two additional putative ADP-RTs, HopO1-1 and HopO1-2. These ADP-RTs possess low similarity with HopU1 and, therefore, probably target other plant proteins. ADP-RTs are well-characterized in animal pathogens; however, the ADP-RT substrates described here are new. Interestingly, IpaH9.8, a type III effector from the animal pathogen *Shigella flexneri*, was recently shown to bind the RRM-containing mammalian splicing factor U2AF[35], resulting in the suppression of pro-inflammatory cytokines[37]. The enzymatic activity of IpaH9.8 is presently unknown. The protein targets of HopU1 and IpaH9.8 indicate that animal and plant pathogens may alter RNA metabolism to quell the eukaryotic immune response.

METHODS SUMMARY

Bacterial strains (and plasmids) and primers used in this work are listed in Supplementary Table 2 and 3, respectively. A Δ*hopU1* mutant was made by homologous recombination. Expression of *hopU1* was assessed with RT–PCR[38]. Construction of *A. thaliana* Col-0 transgenic plants was done using the *Agrobacterium*-mediated floral dip method[39]. DC3000 strains were tested for their ability to elicit an hypersensitive response on *Nicotiana tabacum* cv. Xanthi by infiltrating strains resuspended in 5 mM MES (pH 5.6) at a cell density of 1×10^6 or 1×10^7 cells per ml. Pathogenicity assays with *A. thaliana* plants were performed by spray-inoculation with bacterial suspensions[12]. Seeds of the *Arabidopsis* SALK_039556.21.25.x and SALK_051743.42.85.x lines, corresponding to the *grp7-1* and *grp7-2* mutants, respectively, were obtained from the *Arabidopsis* Biological Resource Center at Ohio State University. *A. thaliana* plants were infiltrated with 1 μM of flg22 and stained with aniline blue for callose as described[25]. Leaves were examined with a Zeiss AxioPlan2 Imaging System microscope with fluorescence. The poly-L-arginine assay was done as previously described[40]. The ADP-RT assay was adapted from a described protocol[41]. GRP7–GST site-directed mutant constructs were made using the QuikChange Site-Directed Mutagenesis kit (Stratagene). *In planta* HopU1 ADP-RT assays with GRP7 were done using *Agrobacterium* delivery as described[21] and assessed with immunoblots.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 February; accepted 8 March 2007.
Published online 22 April 2007.

1. He, S. Y., Nomura, K. & Whittam, T. S. Type III protein secretion mechanism in mammalian and plant pathogens. *Biochim. Biophys. Acta* **1694**, 181–206 (2004).

2. Galán, J. E. & Wolf-Watz, H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **444,** 567–573 (2006).
3. Mota, L. J. & Cornelis, G. R. The bacterial injection kit: type III secretion systems. *Ann. Med.* **37,** 234–249 (2005).
4. Abramovitch, R. B., Anderson, J. C. & Martin, G. B. Bacterial elicitation and evasion of plant innate immunity. *Nature Rev. Mol. Cell Biol.* **7,** 601–611 (2006).
5. Espinosa, A. & Alfano, J. R. Disabling surveillance: bacterial type III secretion system effectors that suppress innate immunity. *Cell. Microbiol.* **6,** 1027–1040 (2004).
6. Mudgett, M. B. New insights to the function of phytopathogenic bacterial type III effectors in plants. *Annu. Rev. Plant Biol.* **56,** 509–531 (2005).
7. Axtell, M. J. & Staskawicz, B. J. Initiation of RPS2-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4. *Cell* **112,** 369–377 (2003).
8. Mackey, D., Belkhadir, Y., Alonso, J. M., Ecker, J. R. & Dangl, J. L. *Arabidopsis* RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell* **112,** 379–389 (2003).
9. Shao, F. *et al.* Cleavage of *Arabidopsis* PBS1 by a bacterial type III effector. *Science* **301,** 1230–1233 (2003).
10. Lopez-Solanilla, E., Bronstein, P. A., Schneider, A. R. & Collmer, A. HopPtoN is a *Pseudomonas syringae* Hrp (type III secretion system) cysteine protease effector that suppresses pathogen-induced necrosis associated with both compatible and incompatible plant interactions. *Mol. Microbiol.* **54,** 353–365 (2004).
11. Bretz, J. R. *et al.* A translocated protein tyrosine phosphatase of *Pseudomonas syringae* pv. *tomato* DC3000 modulates plant defence response to infection. *Mol. Microbiol.* **49,** 389–400 (2003).
12. Espinosa, A., Guo, M., Tam, V. C., Fu, Z. Q. & Alfano, J. R. The *Pseudomonas syringae* type III-secreted protein HopPtoD2 possesses protein tyrosine phosphatase activity and suppresses programmed cell death in plants. *Mol. Microbiol.* **49,** 377–387 (2003).
13. Janjusevic, R., Abramovitch, R. B., Martin, G. B. & Stebbins, C. E. A bacterial inhibitor of host programmed cell death defenses is an E3 ubiquitin ligase. *Science* **311,** 222–226 (2006).
14. Grant, S. R., Fisher, E. J., Chang, J. H., Mole, B. M. & Dangl, J. L. Subterfuge and manipulation: Type III effector proteins of phytopathogenic bacteria. *Annu. Rev. Microbiol.* **60,** 425–449 (2006).
15. Lindeberg, M. *et al.* Closing the circle on the discovery of genes encoding Hrp regulon members and type III secretion system effectors in the genomes of three model *Pseudomonas syringae* strains. *Mol. Plant Microbe Interact.* **19,** 1151–1158 (2006).
16. Guttman, D. S. *et al.* A functional screen for the type III (Hrp) secretome of the plant pathogen *Pseudomonas syringae. Science* **295,** 1722–1726 (2002).
17. Petnicki-Ocwieja, T. *et al.* Genomewide identification of proteins secreted by the Hrp type III protein secretion system of *Pseudomonas syringae* pv. tomato DC3000. *Proc. Natl Acad. Sci. USA* **99,** 7652–7657 (2002).
18. Yates, S. P., Jorgensen, R., Andersen, G. R. & Merrill, A. R. Stealth and mimicry by deadly bacterial toxins. *Trends Biochem. Sci.* **31,** 123–133 (2006).
19. Barbieri, J. T. & Sun, J. *Pseudomonas aeruginosa* ExoS and ExoT. *Rev. Physiol. Biochem. Pharmacol.* **152,** 79–92 (2004).
20. Corda, D. & Di Girolamo, M. Functional aspects of protein mono-ADP-ribosylation. *EMBO J.* **22,** 1953–1958 (2003).
21. Jamir, Y. *et al.* Identification of *Pseudomonas syringae* type III effectors that suppress programmed cell death in plants and yeast. *Plant J.* **37,** 554–565 (2004).
22. Nurnberger, T., Brunner, F., Kemmerling, B. & Piater, L. Innate immunity in plants and animals: striking similarities and obvious differences. *Immunol. Rev.* **198,** 249–266 (2004).
23. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124,** 783–801 (2006).
24. Jones, J. D. & Dangl, J. L. The plant immune system. *Nature* **444,** 323–329 (2006).
25. Gomez-Gomez, L., Felix, G. & Boller, T. A single locus determines sensitivity to bacterial flagellin in *Arabidopsis thaliana. Plant J.* **18,** 277–284 (1999).
26. Whalen, M. C., Innes, R. W., Bent, A. F. & Staskawicz, B. J. Identification of *Pseudomonas syringae* pathogens of *Arabidopsis* and a bacterial locus determining avirulence on both *Arabidopsis* and soybean. *Plant Cell* **3,** 49–59 (1991).
27. Ohta, M., Sugita, M. & Sugiura, M. Three types of nuclear genes encoding chloroplast RNA-binding proteins (cp29, cp31 and cp33) are present in *Arabidopsis thaliana*: presence of cp31 in chloroplasts and its homologue in nuclei/cytoplasms. *Plant Mol. Biol.* **27,** 529–539 (1995).
28. van Nocker, S. & Vierstra, R. D. Two cDNAs from *Arabidopsis thaliana* encode putative RNA binding proteins containing glycine-rich domains. *Plant Mol. Biol.* **21,** 695–699 (1993).
29. Burd, C. G. & Dreyfuss, G. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265,** 615–621 (1994).
30. Jessen, T. H., Oubridge, C., Teo, C.-H., Pritchard, C. & Nagai, K. Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *EMBO J.* **10,** 3447–3456 (1991).
31. Gomez, J. *et al.* A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein. *Nature* **334,** 262–264 (1988).
32. Naqvi, S. M. *et al.* A glycine-rich RNA-binding protein gene is differentially expressed during acute hypersensitive response following Tobacco Mosaic Virus infection in tobacco. *Plant Mol. Biol.* **37,** 571–576 (1998).
33. Heintzen, C., Nater, M., Apel, K. & Staiger, D. AtGRP7, a nuclear RNA-binding protein as a component of a circadian-regulated negative feedback loop in *Arabidopsis thaliana. Proc. Natl Acad. Sci. USA* **94,** 8515–8520 (1997).
34. Staiger, D., Zecca, L., Wieczorek Kirk, D. A., Apel, K. & Eckstein, L. The circadian clock regulated RNA-binding protein *At*GRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. *Plant J.* **33,** 361–371 (2003).
35. Maris, C., Dominguez, C. & Allain, F. H. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272,** 2118–2131 (2005).
36. Alonso, J. M. *et al.* Genome-wide insertional mutagenesis of *Arabidopsis thaliana. Science* **301,** 653–657 (2003).
37. Okuda, J. *et al. Shigella* effector IpaH9.8 binds to a splicing factor $U2AF^{35}$ to modulate host immune responses. *Biochem. Biophys. Res. Commun.* **333,** 531–539 (2005).
38. Guo, M. *et al. Pseudomonas syringae* type III chaperones ShcO1, ShcS1, and ShcS2 facilitate translocation of their cognate effectors and can substitute for each other in the secretion of HopO1-1. *J. Bacteriol.* **187,** 4257–4269 (2005).
39. Bechtold, N., Ellis, J. & Pelletier, G. *In planta Agrobacterium*-mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *C. R. Acad. Sci. Paris Ser. III* **316,** 1194–1199 (1993).
40. Coye, L. H. & Collins, C. M. Identification of SpyA, a novel ADP-ribosyltransferase of *Streptococcus pyogenes*. *Mol. Microbiol.* **54,** 89–98 (2004).
41. Sun, J. & Barbieri, J. T. *Pseudomonas aeruginosa* ExoT ADP-ribosylates CT10 regulator of kinase (Crk) proteins. *J. Biol. Chem.* **278,** 32794–32800 (2003).
42. Han, S. & Tainer, J. A. The ARTT motif and a unified structural understanding of substrate recognition in ADP-ribosylating bacterial toxins and eukaryotic ADP-ribosyltransferases. *Int. J. Med. Microbiol.* **291,** 523–529 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the members of the Alfano laboratory for many fruitful discussions, Y. Zhou and C. Elowsky for technical assistance with confocal microscopy, T. Clemente and S. Sato for constructing transgenic plants, G. Li and C. Bryan for assistance in the identification of the *A. thaliana grp7* mutants, A. Collmer for reviewing the manuscript, P. Seitz for assistance in plasmid constructions, and J. T. Barbieri for help initiating the ADP-RT assays in our laboratory. We are grateful to the Ohio State University *Arabidopsis* Biological Resource Center, the Salk Institute Genomic Analysis Laboratory, and the *Arabidopsis* research community for providing the *Arabidopsis* SALK lines used in this study. This research was supported by grants from the National Science Foundation and the National Institutes of Health, and funds from the Plant Science Initiative at the University of Nebraska to J.R.A, and a grant from the German Research Council to D.S.

Author Contributions Z.Q.F. constructed the DC3000 Δ*hopU1* mutant, made the transgenic HopU1–HA-expressing *A. thaliana* plants, and performed the experiments in Figs 1a, b; 2a, b; 3; 4a, d and Supplementary Figs 1, 2, 4, 5b–d and 6d; M.G. identified the homozygous *A. thaliana grp7* mutant plants, cloned the HopU1–His substrate complementary DNAs, and performed the experiments in Figs 2c, d and 5 and Supplementary Figs 3, 4, 6a–c and 7; B.-r.J. performed the experiments in Figs 1c and 4c and Supplementary Fig. 5a, d; and F.T. provided technical support for several experiments. T.E.E. helped direct the identification of the HopU1–His substrates; R.L.C. performed the mass spectrometry and peptide database searches; D.S. provided the anti-GRP antibody, plasmids pGRP7-Gly and pGRP7-RRM, and insights on RNA-binding proteins; J.R.A. helped design the experimental plan, designed Supplementary Fig. 8, and was the primary writer of the paper. All of the authors discussed the results and commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.R.A. (jalfano2@unl.edu).

METHODS

Bacterial strains and nucleotide primers. Bacterial strains and plasmids used in this work are listed in Supplementary Table 2. The primers used are listed in Supplementary Table 3.

Construction of the DC3000 Δ*hopU1* mutant. A DC3000 Δ*hopU1* mutant was made by amplifying a 2 kb region upstream and downstream of *hopU1* using PCR with primer sets P1058/P1059 and P1078/P1079. The DNA fragment upstream of *hopU1* was ligated into pHP45Ω using *Bam*HI and *Hind*III sites. The DNA fragment downstream of *hopU1* was ligated into the pHP45Ω derivative containing the upstream fragment using *Xba*I and *Sac*I restriction enzymes such that the *hopU1* flanking regions were on either side of an Ω fragment in the same orientation. This cassette was ligated into the broad-host-range vector pRK415 using *Bam*HI and *Sac*I restriction enzymes. The resulting construct was electroporated into DC3000 and homologous recombination was selected for by selecting for retention of the antibiotic marker linked to the mutation, and loss of the plasmid marker. The resulting mutant, UNL141, was confirmed with PCR, using primers that annealed to the flanking regions.

RT–PCR to determine *hopU1* expression. The isolation of DC3000 RNA and reverse transcriptase PCR were performed as previously described[38].

Construction of *A. thaliana* Col-0 transgenic plants expressing HopU1–HA and HopU1$_{DD}$–HA. *hopU1* was amplified with PCR and recombined into the Gateway entry vector pENTR/D-TOPO. Site-directed mutagenesis was carried out on this entry construct to change the two glutamic acids in the ADP-RT active site to aspartic acids using the QuikChange Site-Directed Mutagenesis kit (Stratagene). The resulting constructs were recombined into binary Gateway destination vectors such that the *hopU1* derivatives were downstream of a *CaMV 35S* promoter; these were then electroporated into *Agrobacterium tumefaciens* C58C1 and transformed into *A. thaliana* Col-0 using the *Agrobacterium*-mediated floral dip method[39]. Plants were propagated through four generations and confirmed to constitutively produce HopU1–HA or HopU1$_{DD}$–HA, with immunoblots using anti-HA antibodies.

Plant materials and bioassays. *P. syringae* pv. tomato DC3000 strains were tested for their ability to elicit an hypersensitive response on *Nicotiana tabacum* cv. Xanthi by infiltrating strains resuspended in 5 mM MES (pH 5.6) at a cell density of 1×10^6 or 1×10^7 cells per ml with a needleless syringe. The Δ*hopU1* mutant UNL141 strain carrying plasmids pLN1981 (encoding HopU1), pLN1982 (encoding HopU1$_{DD}$), or pML123 (empty vector) were used in hypersensitive response suppression experiments. Pathogenicity assays with *A. thaliana* Col-0 and *grp7* mutant plants were performed by spray-inoculation with bacterial suspensions that were adjusted to an A_{600} of 0.2, as described[12]. Bacteria were enumerated from leaf tissue by plating dilutions on KB plates with the appropriate antibiotics. Seeds of *Arabidopsis* SALK lines SALK_039556.21.25.x and SALK_051743.42.85.x, which were used to identify *grp7-1* and *grp7-2* mutants, respectively, were obtained from the Ohio State University Biological Resource Center. Further description of methods used to confirm the homozygosity of the T-DNA inserts and the absence of *GRP7* RNA and GRP7 protein is in the Supplementary Methods.

Microscopic detection of callose deposition. *A. thaliana* Col-0 wild-type, Col-0 *grp7* mutant plants and *A. thaliana* expressing HopU1–HA or HopU1$_{DD}$–HA were infiltrated with 1 μM of flg22. After 16 h, the leaves were excised, cleared, and stained with aniline blue for callose, as described[25]. Leaves were examined with a Zeiss AxioPlan2 Imaging System microscope with fluorescence. The number of callose depositions was determined with Quantity One software (Bio-Rad). Twenty fields of view (each 0.56 mm^2) were analysed and averaged. The average and standard errors of three independent assays for each treatment were recorded.

Purification of HopU1–His and HopU1$_{DD}$–His mutation derivative. HopU1-His and HopU1$_{DD}$-His mutation derivative were affinity-purified from *Escherichia coli* BL21 (DE3) using Ni-NTA resin following the manufacturer's instructions (Qiagen). Purified HopU1–His and HopU1$_{DD}$–His were stored in aliquots at −80 °C or on ice. HopU1–His and HopU1$_{DD}$–His protein concentrations were measured with the BioRad Protein Assay Kit and the purity of the proteins was examined by SDS–polyacrylamide gel electrophoresis (PAGE).

Poly-L-Arginine ADP-RT assay. To determine if HopU1 ADP-ribosylated the artificial substrate poly-L-arginine, we used a previously described protocol[40]. Briefly, homo-poly-L-arginine (80 μl of 10 mg ml^{-1} in 0.1 M dimethyl glutaric acid buffer pH 7.0), 1 μM [^{32}P]-NAD (GE Healthcare) radiolabelled on the ADP-ribose moiety and 1 μM HopU1–His or HopU1$_{DD}$–His were incubated at room temperature for 1 h, suspended in 0.1 M phosphate buffer, centrifuged, and resuspended in 250 μl of 0.1 M HCl and 500 μl of 0.1 M dimethyl glutaric acid buffer (pH 7.0). Incorporated radioactivity was determined using liquid scintillation. BSA (1 μM) was used as a control.

ADP-RT assays with plant extracts and recombinant substrates. The ADP-RT assay was adapted from a described protocol[41]. Briefly, approximately 10 ng of partially-purified HopU1–His was incubated with 5 μCi of [^{32}P]-NAD (1 Ci μmole^{-1}) and plant extracts (*A. thaliana* Col-0 or tobacco), *E. coli* extracts, or 1–2 μg of partially purified recombinant substrates. The reaction was stopped by adding SDS sample buffer, after incubation at room temperature for 1 h. Samples were subjected to one-dimensional or two-dimensional SDS–PAGE followed by autoradiography of dried gels.

***GRP7–GST* plasmids and site-directed mutagenesis of *GRP7–GST*.** The nucleotides corresponding to the RRM domain and the glycine-rich domain of GRP7 were separately PCR-amplified using primer sets P2578/P2579 and P2580/P2581, respectively, and ligated into the GST vector pGEX-5X-1. The resulting plasmids encoded an RRM–GST fusion (pGRP7-RRM) and glycine-rich-domain–GST fusion (pGRP7-Gly). Several constructs encoding full-length GRP7–GST fusions in which individual arginine residues in the RRM domain were substituted with lysines were made using the QuikChange Site-Directed Mutagenesis (Stratagene). The primers used to introduce these mutations and the resulting plasmids are listed in Supplementary Tables 2 and 3. These constructs were separately expressed in *E. coli* to make lysates that contained GRP7–GST derivatives, which were used in ADP-RT assays.

***In planta* HopU1 ADP-RT assays with GRP7, using *Agrobacterium* delivery.** *Agrobacterium* transient co-delivery of *GRP7–HA* and *hopU1–Flag* or *GRP7–HA* and *hopU1$_{DD}$–Flag* DNA were done as described[21]. After 40 h, plant tissue was harvested with a No. 8 cork borer (12-mm diameter), ground in liquid nitrogen, and resuspended in SDS sample buffer. Samples were subjected to SDS–PAGE and immunoblot analysis using anti-Flag or anti-HA antibodies to determine the molecular mass of GRP7 in the presence or absence of active HopU1–Flag.

LETTERS

Shear heating as the origin of the plumes and heat flux on Enceladus

F. Nimmo[1], J. R. Spencer[2], R. T. Pappalardo[3] & M. E. Mullen[4]

Enceladus, a small icy satellite of Saturn, has active plumes jetting from localized fractures ('tiger stripes') within an area of high heat flux near the south pole[1–4]. The plume characteristics[1] and local high heat flux[2] have been ascribed either to the presence of liquid water within a few tens of metres of the surface[1], or the decomposition of clathrates[5]. Neither model addresses how delivery of internal heat to the near-surface is sustained. Here we show that the most likely explanation for the heat[2] and vapour production[6,7] is shear heating by tidally driven lateral (strike-slip) fault motion[1,8,9] with displacement of ~0.5 m over a tidal period. Vapour produced by this heating may escape as plumes through cracks reopened by the tidal stresses[10]. The ice shell thickness needed to produce the observed heat flux is at least 5 km. The tidal displacements required imply a Love number of $h_2 > 0.01$, suggesting that the ice shell is decoupled from the silicate interior by a subsurface ocean. We predict that the tiger-stripe regions with highest relative temperatures will be the lower-latitude branch of Damascus, Cairo around 60° W longitude and Alexandria around 150° W longitude.

Mid-infrared observations of the tiger stripes show spatially varying temperature anomalies with mean temperatures a few kelvin higher than the background, and also require regions a few hundred metres wide with temperatures up to at least 150 K (ref. 2). The total heat flow for the region south of 65° S is 3.9–7.7 GW (ref. 2). Stellar occultation and mass spectrometer data suggest that the vapour plume mass flux is 120–180 kg s^{-1} (refs 6, 7). The mass of solid ice grains relative to water vapour in the plume has been used to argue that the plume is generated by shallow boiling liquid water[1]. Alternatively, clathrate decomposition may be able to produce the observed plume[5]. However, neither of these models addresses how heat is transported to the near-surface to sustain these putative processes.

Lateral (strike-slip) motion has previously been inferred for Jupiter's moon Europa[11], which probably possesses a subsurface ocean[12]. Such motion would occur as a result of the time-varying orientation of tidal stresses on that body[13], and may result in localized heating due to friction or viscous dissipation[8]. Faults in the polar regions of Enceladus will experience stresses driving strike-slip motion for part of each tidal cycle[13,14] (see Supplementary Information). According to ref. 15, the mean shear velocity on an individual fault driven by these tidal stresses is given by $u = (\mathrm{d}\varepsilon/\mathrm{d}t)d \approx (4 \times 10^{-5} h_2)$ m s^{-1} (see Supplementary Information). Here d is the distance between faults (assumed to be 30 km for the tiger stripes[1]), h_2 is a dimensionless (Love) number that depends on the satellite's internal structure, and $\mathrm{d}\varepsilon/\mathrm{d}t$ is the mean diurnal tidal strain rate. For a uniform body that lacks significant rigidity (h_2 approaching 2.5) and fault friction, the shear velocity will be of the order of 10^{-4} m s^{-1}. The shear velocity controls the shear heating rate. The presence of fault friction, an elastic ice shell, or a silicate interior will reduce h_2 and the shear velocity by as much as several orders of magnitude[16] (see below).

To calculate the thermal consequences of this strike-slip motion, we use an approach[8,9] in which a single shear zone generates local heating in a floating ice shell. We adopt a nominal conductive ice layer thickness of 24 km (compare with ref. 17), comparable to the tiger-stripe spacing, which may be underlain by a convecting icy interior[18] and a global subsurface ocean or a localized sea. The shearing is accomplished by motion on a discrete vertical fault surface at shallow depths where the ice is cold and brittle, and by ductile motion across a broader shear zone at greater depths. The transition depth between brittle and ductile deformation is calculated in a self-consistent manner[9], and the resulting equilibrium temperature structure is derived. The surface temperature is calculated based on the subsurface and radiative heat fluxes (see Supplementary Information).

The energy generated by shearing will be partitioned principally between heating and vapour sublimation[19]. We adopt a dimensionless free parameter α to describe this partitioning, where $\alpha = 0$ implies that all the energy goes into vapour production, and $\alpha = 1$ means all the energy goes into heating. To provide a zeroth-order description of the likely vapour transport processes (see Supplementary Information) we solve the diffusion equation[20,21] $\partial\rho_v/\partial t = D\nabla^2\rho_v + (1-\alpha)H/L\phi$ to the steady state, where ρ_v is the pore vapour density, D and L are the (temperature-dependent) diffusivity[20] and latent heat of sublimation[22], respectively, H is the total volumetric heat production in the ice calculated for the steady-state temperature distribution, and ϕ is the porosity[21]. In steady-state the vapour escape rate is independent of the permeability structure assumed (see Supplementary Information). Although we do not consider the timing or details of volatile release, vapour escape is probably mainly through the brittle shear zone that is opened and closed by the periodic tidal stresses[10].

Figure 1 shows that the temperatures and temperature gradients near the strike-slip zone are significantly increased as a result of the brittle and viscous heating when the shear velocity is 8×10^{-6} m s^{-1} (~0.5 m slip per tidal cycle; $h_2 \approx 0.2$) and $\alpha = 0.1$. The brittle layer thickness is locally reduced to 4 km along the fracture, implying that peak shear stresses of ~100 kPa are required to cause motion, comparable to the tidal shear stresses expected for $h_2 \approx 0.2$ (see Supplementary Information).

Because the bulk of the energy is partitioned into vapour production, the total conductive heat flow (0.95 GW for a total fault length of 500 km; ref. 1) and maximum surface temperature increase (1.1 K) are smaller than the infrared observations[2]. On the other hand, the vapour production rate of ~2,000 kg s^{-1} is ten times larger than the inferred plume flux[7]. We propose that the infrared signature is the result of latent heat released by local recondensation of ~90% of the vapour in cold, near-surface porous ice around the shear zone[5],

[1]Department of Earth and Planetary Sciences, University of California Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. [2]Department of Space Studies, Southwest Research Institute, 1050 Walnut Street, Boulder, Colorado 80304, USA. [3]Planetary Science and Life Detection Section, Jet Propulsion Laboratory, California Institute of Technology, Mail Stop 183-301, Pasadena, California 91109, USA. [4]Department of Astrophysical and Planetary Sciences, Campus Box 391, University of Colorado, Boulder, Colorado 80309-0391, USA.

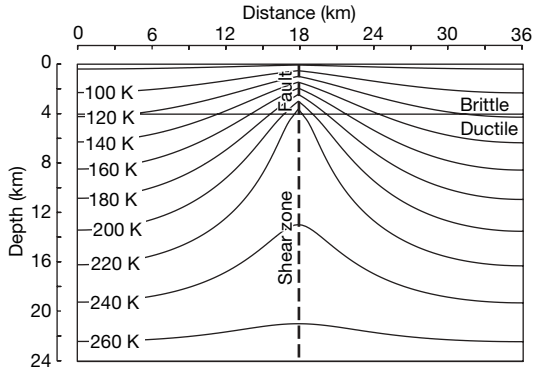


Figure 1 | Shear heating model. Cross-section of steady-state temperature structure (in kelvin) for a shear velocity of 8×10^{-6} m s^{-1} (implying a Love number $h_2 \approx 0.2$; see text) and $\alpha = 0.1$ (implying that 90% of the dissipation goes into vapour production rather than temperature increases). The bold dashed line represents a vertical fault along which lateral shear occurs (perpendicular to the page); the brittle layer thickness at the shear zone is calculated self-consistently[9]. This model produces 0.95 GW of conductive heat flow (assuming a 500 km length) and a total vapour production rate of 2,160 kg s^{-1}; recondensation of the vapour would add a further 6 GW of heat. The ice viscosity is temperature-dependent (ref. 27, see Supplementary Information). Calculations assume coefficient of friction $f = 0.3$, $g = 0.11$ m s^{-2}, ice density $\rho = 930$ kg m^{-3}, ice reference viscosity 10^{13} Pa s, thermal conductivity 3 W m^{-1} K^{-1}, grid spacing 0.3 km (horizontal) and 0.2 km (vertical), surface temperature in absence of subsurface heating 75 K (ref. 2); other parameters identical to ref. 8. The results are relatively insensitive to changes in viscosity or shell thickness (see Supplementary Information).

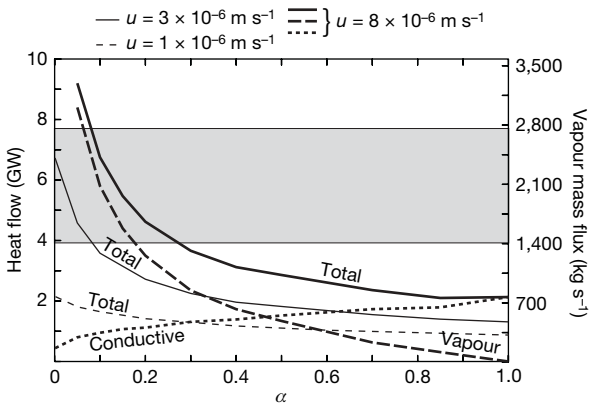


Figure 2 | Heat flow and vapour flux as a function of shear velocity *u* and parameter α. The free parameter α controls what fraction of energy dissipated goes into vapour production. The vapour curve refers to both vapour mass flux and heat flow; other curves refer to heat flow only. Vapour production is calculated using the Methods described in the Supplementary Information. Vapour-driven heat flow is calculated by assuming that all the vapour recondenses with a latent heat of 2.8 MJ kg^{-1}. Conductive heat flow is obtained by integrating the surface heat flux across the computational domain. Total tiger-stripe length is assumed to be 500 km and shell thickness is 24 km. Increased vapour production produces a colder shear zone and allows the brittle layer to extend to greater depths, resulting in an overall increase in heat production. The shaded zone indicates the surface heat flow deduced from Cassini observations[2]. With a 24-km-thick shell, a shear velocity of 10^{-6} m s^{-1} ($h_2 = 0.02$) is insufficient to match the observations. A conservatively thick (81 km) shell requires $h_2 > 0.01$.

while the remaining vapour escapes to form the observed plumes. Recondensation of 1,800 kg s^{-1} vapour over a shallow subsurface zone 500 m wide and 500 km long results in a heat flow of 5 GW and an equivalent surface temperature of 137 K, consistent with the infrared observations[2]. Heating and consequent ice grain growth due to this recondensation is also a plausible explanation of the greater inferred grain size at the tiger stripes[1,23].

The modelled flux-weighted temperature of the vapour entering the shear zone is ~175 K (root-mean-square thermal velocity ~500 m s^{-1}). Particles escaping the shear zone upwards are likely to be collimated, resulting in a vertical velocity comparable to the thermal velocity, but much larger than the horizontal velocity. The thermal velocity we obtain is thus consistent with modelling of the Cassini vapour observations[7]. The Cassini-derived particle ejection rate[1] was about 1/40th of the estimated 150 kg s^{-1} gas production rate[7] observed four months earlier. Because most of the ~2,000 kg s^{-1} of vapour generated in our model is assumed to condense near the surface, the entrainment in the plume of a small fraction (<1%) of the condensing solids is plausible without invoking near-surface liquid water.

Figure 2 shows that our results are sensitive to the shear velocity and value for α adopted. Greater vapour production (smaller α) results in less heating of the ice, an increased depth of brittle deformation, and a greater overall power output. A shear velocity of 10^{-6} m s^{-1} results in a total power output less than that of the

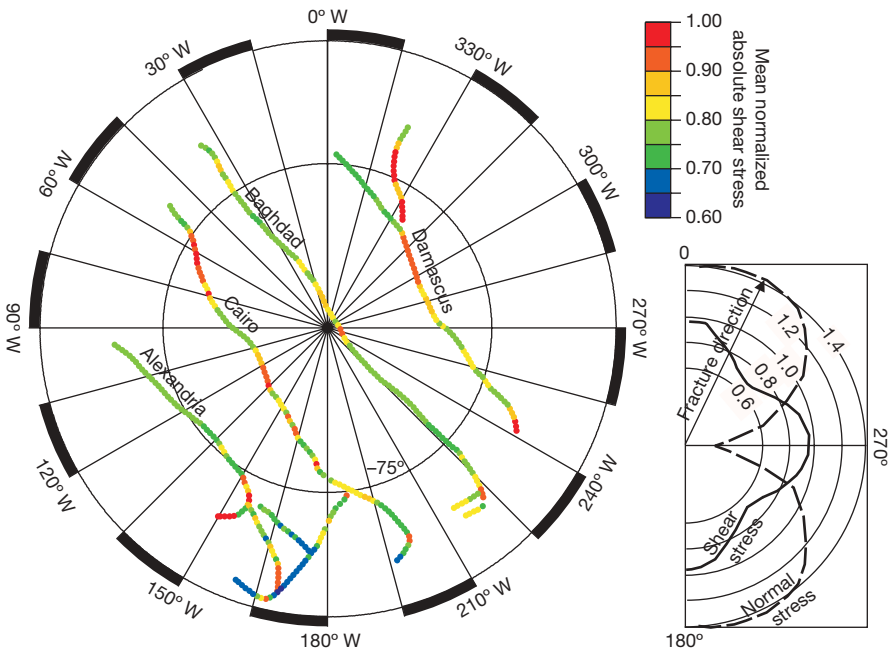


Figure 3 | Predicted tiger-stripe stresses. Plotted is a polar stereographic projection of Enceladus tiger stripes and related structures south of 64° S. Colours indicate mean absolute shear stress (averaged over one tidal cycle), resolved onto the local tiger-stripe orientation and normalized to the maximum value (see Supplementary Information for further details). Areas with higher resolved shear stresses will undergo greater long-term shear heating and vapour production. For brittle heating, the long-term heat production rate will be linearly proportional to the plotted shear stress. Thus, areas with higher stress will show higher temperatures. Cairo is a particularly good target for future Cassini observations because of the predicted variations in temperature along its length. The inset plots the normalized resolved shear and normal stresses as a function of fracture orientation for a fracture located at 70° S, 165° W. Circular contours give normalized stress magnitudes. Normal stresses are maximized on fractures oriented close to the direction of 0° or 180°; shear stresses are maximized on fractures oriented in the direction of 0°, 90°, 180° or 270°.

infrared observations. For a conservatively thick (81 km) ice shell, the same constraint results in a lower bound on h_2 of 0.01.

We have simplified or neglected many details of the vapour sublimation and transport mechanisms, such as their effect on the temperature[19,24] and porosity structure[21], and the potential for time-variable transport and release[6,10]. Nonetheless, Fig. 2 shows that shear heating can generate enough power and vapour to account for the observations. For the shell properties we assume, shell thicknesses in excess of 5 km are required to match the observed heat flow (Supplementary Information).

Different tiger stripes will experience different rates of long-term heating, due to the spatial variation in tidal stresses and the tiger stripe orientation. Figure 3 plots the time-averaged absolute shear stress resolved onto the local tiger stripe orientation and normalized to the maximum shear stress (Supplementary Information). The shear velocity, and thus the long-term heating rate, will be proportional to this quantity if brittle heating dominates[8,9]. Despite their variable spatial resolution, some existing thermal infrared observations[2] support our model. In particular, the two branches of Damascus show a temperature variation consistent with the prediction, and Cairo has a hotspot at about 120° longitude, as predicted. On the other hand, Baghdad temperatures are higher than expected.

The inset in Fig. 3 shows how the resolved stresses vary as a function of fracture azimuth for near-polar fractures. The current orientation of the tiger stripes results in near-maximum normal stresses, suggesting they may have initiated as tensile cracks from contemporary diurnal tides.

In our shear-heating model, the shear velocities required to match the observations imply a Love number $h_2 > 0.01$. A global subsurface ocean decouples the ice shell from the silicate interior and permits larger shear velocities and greater tidal heating[16,17]. Figure 4 shows the Love number h_2 as a function of ice shell viscosity for spherically symmetric models with and without a subsurface ocean. A model with a global subsurface ocean can reproduce the shear velocities required for reasonable assumptions of ice viscosity. The shear heating model is also consistent with reorientation of the south polar region owing to density anomalies in a relatively thick ice shell; the presence of an ocean argues against that density anomaly being located in a decoupled rocky core[25]. In the absence of an ocean, unrealistically low ice shell (or silicate mantle) viscosities are required. To generate the shear velocities required, it seems likely that Enceladus has a subsurface ocean, either global or possibly localized[26], beneath a thick ice shell.

Our hypothesis suggests that permanent strike-slip offsets might be present[13] and potentially observable, although the cyclical nature of the shearing could result in negligible net strike-slip offset. More important, the regions with the highest relative temperatures are predicted to be the lower-latitude branch of Damascus, Cairo around 60° W longitude, and Alexandria around 150° W longitude (Fig. 3).

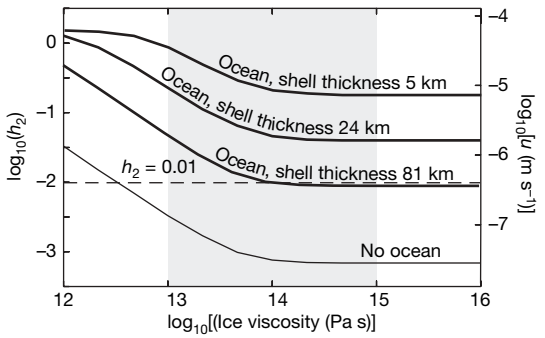


Figure 4 | Effect of interior structure on mean shear velocity. Love numbers h_2 and resulting mean shear velocities u ($= 4 \times 10^{-5} h_2$ m s^{-1}; see text) for models of a spherically symmetric Enceladus with and without an interior ocean are plotted as a function of ice shell viscosity. The shaded region denotes the probable range in viscosity for a convecting ice shell with a grain size of 0.1–10 mm and a temperature of 250–270 K (refs 27, 28); the dashed line plots the estimated lower bound on h_2 (see text). Calculations were carried out using the method of ref. 16 and using parameters similar to those in ref. 25. The silicate core has a radius of 161 km, density of 3,500 kg m^{-3}, a rigidity of 100 GPa and viscosity of 10^{21} Pa s. The ocean (where present) plus ice shell have a combined thickness of 91 km. The ice shell has a density and rigidity of 950 kg m^{-3} and 3 GPa, respectively, a constant assumed viscosity, and a thickness of 5, 24 or 81 km. The silicate mantle viscosity would have to be reduced to an unrealistically low 10^{14} Pa s or less to allow any significant increase in deformation of the model lacking an ocean. Including the effect of the conductive ice lid would further reduce the value of h_2. For large ice viscosities the body behaves in a purely elastic fashion and h_2 is independent of viscosity.

Received 8 December 2006; accepted 22 March 2007.

1. Porco, C. C. *et al.* Cassini observes the active south pole of Enceladus. *Science* **311,** 1393–1401 (2006).
2. Spencer, J. R. *et al.* Cassini encounters Enceladus: background and the discovery of a south polar hot spot. *Science* **311,** 1401–1405 (2006).
3. Spahn, F. *et al.* Cassini dust measurements at Enceladus and implications for the origin of the E ring. *Science* **311,** 1416–1418 (2006).
4. Hansen, C. J. *et al.* Enceladus' water vapour plume. *Science* **311,** 1422–1425 (2006).
5. Kieffer, S. W. *et al.* A clathrate reservoir hypothesis for Enceladus' south polar plume. *Science* **314,** 1764–1766 (2006).
6. Waite, J. H. *et al.* Cassini ion and neutral mass spectrometer: Enceladus plume composition and structure. *Science* **311,** 1419–1422 (2006).
7. Tian, F., Stewart, A. I. F., Toon, O. B., Larsen, K. M. & Esposito, L. W. Monte Carlo simulations of the water vapor plumes on Enceladus. *Icarus* (in the press).
8. Nimmo, F. & Gaidos, E. Thermal consequences of strike-slip motion on Europa. *J. Geophys. Res.* **107,** 5021, doi:10.1029/2000JE001476 (2002).
9. Prockter, L. M., Nimmo, F. & Pappalardo, R. T. A shear heating origin for ridges on Triton. *Geophys. Res. Lett.* **32,** L14202, doi:10.1029/2005GL022832 (2005).
10. Hurford, T. A., Helfenstein, P., Hoppa, G. V., Greenberg, R. & Bills, B. G. Eruptions arising from tidally controlled periodic openings of rifts on Enceladus. *Nature* doi:10.1038/nature05821 (this issue).
11. Hoppa, G. *et al.* Distribution of strike-slip faults on Europa. *J. Geophys. Res.* **105,** 22617–22627 (2000).
12. Zimmer, C., Khurana, K. K. & Kivelson, M. G. Subsurface oceans on Europa and Callisto: constraints from Galileo magnetometer observations. *Icarus* **147,** 329–347 (2000).
13. Hoppa, G., Tufts, B. R., Greenberg, R. & Geissler, P. Strike-slip faults on Europa: global shear patterns driven by tidal stress. *Icarus* **141,** 287–298 (1999).
14. Helfenstein, P. & Parmentier, E. M. Patterns of fracture and tidal stresses on Europa. *Icarus* **53,** 415–430 (1983).
15. Greenberg, R. *et al.* Tectonic processes on Europa: tidal stresses, mechanical response and visible features. *Icarus* **135,** 64–78 (1998).
16. Moore, W. B. & Schubert, G. The tidal response of Europa. *Icarus* **147,** 317–319 (2000).
17. Ross, M. N. & Schubert, G. Viscoelastic models of tidal heating on Enceladus. *Icarus* **78,** 90–101 (1989).
18. Squyres, S. W., Reynolds, R. T., Cassen, P. M. & Peale, S. J. The evolution of Enceladus. *Icarus* **53,** 319–331 (1983).
19. Komle, N. I. *et al.* Temperature evolution and vapour pressure build-up in porous ices. *Planet. Space Sci.* **40,** 1311–1323 (1992).
20. Clifford, S. M. & Hillel, D. The stability of ground ice in the equatorial region of Mars. *J. Geophys. Res.* **88,** 2456–2474 (1983).
21. Schorghofer, N. & Aharonson, O. Stability and exchange of subsurface ice on Mars. *J. Geophys. Res.* **110,** E05003, doi:10.1029/2004JE002350 (2005).
22. Cowan, J. J. & A'Hearn, M. F. Vaporization of comet nuclei: light curves and life times. *Moon Planets* **21,** 155–171 (1979).
23. Brown, R. H. *et al.* Compsition and physical properties of Enceladus' surface. *Science* **311,** 1425–1428 (2006).
24. Espinasse, S., Klinger, J., Ritz, C. & Schmitt, B. Modeling of the thermal behaviour and of the chemical differentiation of cometary nuclei. *Icarus* **92,** 350–365 (1991).
25. Nimmo, F. & Pappalardo, R. T. Diapir-induced reorientation of Saturn's moon Enceladus. *Nature* **441,** 614–616 (2006).
26. Collins, G. C. & Goodman, J. C. Enceladus's south polar sea. *Icarus* (in the press).
27. Goldsby, D. L. & Kohlstedt, D. L. Superplastic deformation of ice: Experimental observations. *J. Geophys. Res.* **106,** 11017–11030 (2001).
28. Pappalardo, R. T. *et al.* Geological evidence for solid-state convection in Europa's ice shell. *Nature* **391,** 365–368 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Gleeson and Z. Crawford for their efforts. This work was funded by NASA's Planetary Geology and Geophysics and Outer Planets Research programmes.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.N. (fnimmo@es.ucsc.edu).

LETTERS

Eruptions arising from tidally controlled periodic openings of rifts on Enceladus

T. A. Hurford[1], P. Helfenstein[2], G. V. Hoppa[3], R. Greenberg[4] & B. G. Bills[1,5]

In 2005, plumes were detected near the south polar region of Enceladus[1], a small icy satellite of Saturn. Observations of the south pole revealed large rifts in the crust, informally called 'tiger stripes', which exhibit higher temperatures than the surrounding terrain and are probably sources of the observed eruptions[2]. Models of the ultimate interior source for the eruptions are under consideration[1,3–5]. Other models of an expanding plume[6] require eruptions from discrete sources, as well as less voluminous eruptions from a more extended source, to match the observations. No physical mechanism that matches the observations has been identified to control these eruptions. Here we report a mechanism in which temporal variations in tidal stress open and close the tiger-stripe rifts, governing the timing of eruptions. During each orbit, every portion of each tiger stripe rift spends about half the time in tension, which allows the rift to open, exposing volatiles, and allowing eruptions. In a complementary process, periodic shear stress along the rifts also generates heat along their lengths[7–9], which has the capacity to enhance eruptions. Plume activity is expected to vary periodically, affecting the injection of material into Saturn's E ring[10] and its formation, evolution and structure. Moreover, the stresses controlling eruptions imply that Enceladus' icy shell behaves as a thin elastic layer, perhaps only a few tens of kilometres thick.

Enceladus is a small (radius ~250 km) moon that orbits Saturn between the moons Mimas and Tethys with a period of 1.37 days. This proximity to Saturn means that tidal dissipation should have quickly circularized the orbit. However, a 2:1 mean motion resonance with the moon Dione, which orbits just beyond Tethys, excites the orbital eccentricity to the observed value of 0.0047, which in turn produces periodic tidal stress on the surface.

A thin-shell approximation yields the surface stresses from the tidal deformation of a thin, elastic outer layer[11], assuming that this layer is decoupled from the deeper interior of the body (for example, by a fluid layer[12,13]). Thus there is negligible shear between the elastic shell and the interior. The thin-shell approximation of surface stress is valid as long as the elastic outer layer is not thick enough to begin significantly hampering tidal deformation of the body. For Europa, the stresses in an icy elastic outer layer of $<$15 km have been shown not to differ significantly from the thin-shell approximation of surface stress[14]. The thin elastic layer deforms to fit the tidal figure of the body's interior. The horizontal strain of this shell, as required to fit the tidally distorted interior, produces stress on the surface, given by:

$$\sigma_{\theta\theta} = (3M\mu h_2/8\pi\rho_{av}a^3)(1+\nu)/(5+\nu)(5+3\cos 2\theta) \quad (1)$$

and

$$\sigma_{\varphi\varphi} = -(3M\mu h_2/8\pi\rho_{av}a^3)(1+\nu)/(5+\nu)(1-9\cos 2\theta) \quad (2)$$

where θ is the angular distance measured with respect to the tidal distortion's axis of symmetry (ostensibly the sub-anti-saturnian line)[15,16]. The quantity $\sigma_{\theta\theta}$ is the stress along the surface in the direction radial to the tidal bulges, while $\sigma_{\varphi\varphi}$ is the stress along the surface in a direction orthogonal to the $\sigma_{\theta\theta}$ stress. In these expressions μ is the rigidity of the shell, ν is Poisson's ratio, ρ_{av} is the average density of the body, and M is the mass of Saturn, the tide-raiser, which is a distance a from the centre of the deformed body. The magnitude of the tidal stress on the surface is proportional to the Love number h_2, which is a constant of proportionality that governs the tidal deformation of the body, and is determined by its internal structure.

Enceladus' finite eccentricity causes small daily changes in a and θ. For example, during Enceladus' orbit, between pericentre and apocentre, the change in distance from Saturn produces a small variation in the height of the main tide with an amplitude of $(9eh_2MR)/(4\rho_{av}a^3)$, where R is Enceladus' radius and e is its eccentricity. In addition to the change in the amplitude of the tide, Saturn's angular position with respect to a fixed location above Enceladus' surface oscillates with an amplitude of $2e$ radians east and west, changing

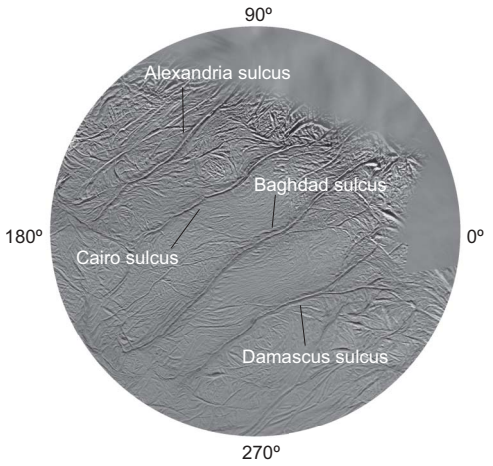


Figure 1 | The south polar region of Enceladus. This image from a high-resolution atlas[21] of Enceladus shows the south polar region from 65° S poleward. These features are typically about 130 km in length, 2 km wide, with a trough 500 m deep, and are flanked on each side by ridges 100 m tall[1]. One of these features (Cairo sulcus) has arcuate segments similar to the cycloidal ridges on Europa[22], where distinctive cycloidal patterns were probably produced by periodic tidal stresses[14,23]. The similar shape suggests a similar formation process controlled by diurnal tidal stresses[22].

[1]Planetary Geodynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. [2]CRSR, Cornell University, Ithaca, New York 14853, USA. [3]Raytheon, Woburn, Massachusetts 01801, USA. [4]Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA. [5]Institute for Geophysics and Planetary Physics, Scripps Institution of Oceanography, La Jolla, California 92093, USA.

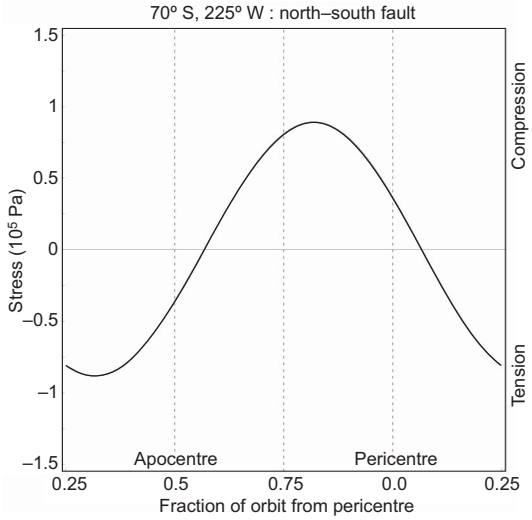


Figure 2 | Stress normal to a north–south-oriented fault. The stress across this fault, which is oriented north–south at a latitude of 70° S and longitude of 225° W is shown. At pericentre the fault is being held closed, because the stress normal to it is compressive. However, as Enceladus orbits, the stress changes from compression to tension, allowing the fault to open, possibly exposing liquid water or other volatiles and creating an eruption. The fault remains in tension for half of an orbit, at which point the normal stress once again becomes compressive, forcing the crack to close, ending any possibility of an eruption.

the angular distance to the tidal bulge θ slightly. These two effects combine, yielding the diurnally varying part of the tide.

For a water-ice crust, we used the plausible values of the elastic parameters[17] $\mu = 3.52 \times 10^9$ Pa and $\nu = 0.33$, and assumed a conservative value for the tidal response given by $h_2 = 0.32$, which corresponds to a diurnal tidal amplitude of ~5 m when $3M/8\rho_{av}a^3 = 0.01$.

In a method similar to that used to consider stress along cracks on the moon Europa's surface[18], diurnal variation in the surface stress normal to a fault can be calculated (Fig. 2). Tension across a fault opens the feature, allowing pent-up volatiles to be released and eruptions to occur. The time of this release depends on the orientation of the fault and its location. Similar analysis of the stresses normal to the fault can be made all along each tiger-stripe feature near the south pole and comparisons with plume observations can be done to see if any of the features are in tension during the observational sequences.

It was hypothesized that similar venting might be observed on Europa, so the Galileo spacecraft imaging team included a sequence specifically designed to search for plumes. However, at the time and place of those observations, the area imaged was under tidal compression—the least likely condition for venting—and no plumes were detected[19]. In view of this, we can address whether the stress state of the cracks on Enceladus was conducive to venting at the times of the Cassini imaging.

The plume activity on Enceladus was observed at different locations in its orbit during imaging sequences that lasted about an hour. During the January observation, Enceladus was about an eighth of an

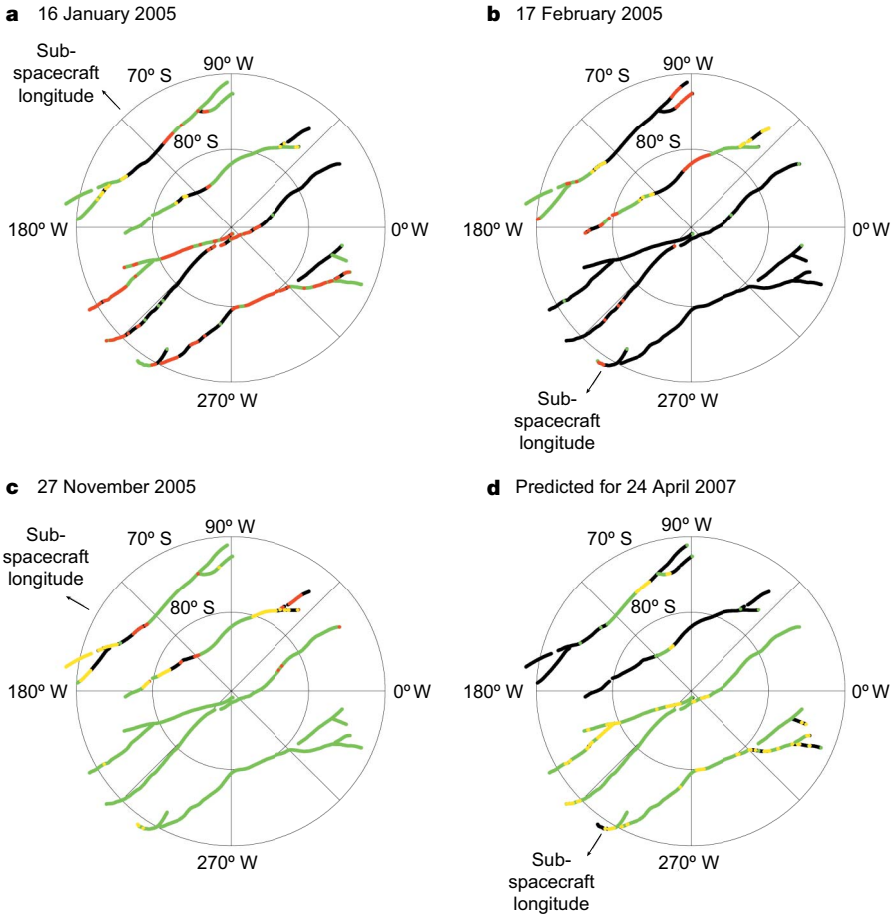


Figure 3 | The stress state across the tiger stripe rifts. During observation sequences from 2005 the stress along each rift is shown (**a**–**c**), along with the predicted state during the proposed April 2007 observation (**d**). The colour indicates the stress state. Black means that some portion of the feature was in compression during the entire imaging sequence, while green means that it was in tension. Yellow indicates that the stress switched from compression to tension, opening the rift during the imaging sequence, and orange indicates that the stress switched from tension to compression, potentially closing the rift.

orbit past apocentre and 46% of the tiger stripes were in tension for the hour-long imaging sequence (Fig. 3a). Tiger stripes between 225° W and 45° W longitude (11% of the rift system) experienced a switch to compression. This observation was scheduled during the part of the orbit when the rifts were in the process of closing as the stresses became compressive. Hence, during the February (Fig. 3b) observations, taken at about at an eighth of an orbit before pericentre, only 16% of the tiger stripes remained in tension while more of the system (4%) switched to compression. We would therefore expect the plume activity in the February detection to be less intense than in the plumes of January, but the different viewing geometries make direct comparisons between the observations difficult, because particles in the plumes are strongly forward-scattering and the intensity of scattered light from the plumes dramatically increases with increasing phase angle. Finally, during the November observation sequence, taken near apocentre (Fig. 3c), 82% of the tiger stripes experienced tension during the imaging sequence.

The three observations of plumes near the south pole show a progression as Enceladus orbits from apocentre towards pericentre. In November, with Enceladus near apocentre, almost the entire region is experiencing tension along the rifts (Fig. 3c). In January, the rifts are beginning to close after passing apocentre, beginning with Damascus sulcus and Baghdad sulcus (Fig. 3a). As Enceladus approaches pericentre, most of the system is in compression (Fig. 3b). We predict that activity should be lowest when Enceladus is near pericentre and highest when Enceladus is near apocentre (see the Supplementary Information).

Because most of the system is in tension during the November observations, jets can be observed erupting nearly anywhere along the rifts (see Supplementary Information). However, if tides control the eruption process, we predict that improved triangulation methods[20] will find that the observed jets are not distributed evenly throughout the rift system. The location of jets may be controlled by tides in two different ways. First, if jets occur where a rift first opens in tension, allowing the pent-up volatiles to erupt, favoured recent jet locations will be found along Alexandria sulcus near 160° W, 72° S, along Cairo sulcus near 170° W, 77° S and 50° W, 75° S, along Baghdad sulcus near 210° W, 70° S and along Damascus sulcus near 240° W, 70° S. Second, tidal stress levels may influence the locations where jetting occurs. The locations along each rift where tidal stresses are close to their maxima are near 105° W, 74° S along Alexandria sulcus, near 115° W, 84° S along Cairo sulcus, near 315° W, 89° S along Baghdad sulcus and near 275° W, 80° S along Damascus sulcus. Thus, of these locations, the ones in Baghdad sulcus and Damascus sulcus opened earlier than the locations along the other two rifts and Baghdad sulcus or Damascus sulcus or both are the most likely sources for the November jets.

A Cassini imaging sequence now planned for 24 April 2007 would take place over about 1.5 h, while Enceladus is about one-fifth of an orbit past pericentre (Fig. 3d). We find that during this time 57% of the tiger stripes will be in tension. Like the November plume observation, a significant portion (6%) of the features will be in the process of opening up as the tension across them changes from compression to tension.

On the basis of the geometry of the imaging sequence in April, sources for plumes erupting from the region of Baghdad sulcus, Alexandria sulcus and Cairo sulcus will be difficult to locate on the surface. Without the surface being resolved in the image a plume in that region could originate from Baghdad sulcus in the foreground or Alexandria sulcus in the background. However, activity from Damascus sulcus should be observationally distinct from the rest of the tiger-stripe rift system and the rift will have recently experienced tension, providing an opportunity to observe the beginning stages of eruptions on Enceladus and test the two scenarios described above. If eruptions are controlled by the release of pent-up volatiles, this observational sequence could reveal an especially active phase as pent-up volatiles begin to erupt along Damascus. However, if stresses near their maxima control the formation of jets then this rift may be observed to be inactive.

Received 29 November 2006; accepted 30 March 2007.

1. Porco, C. C. *et al.* Cassini observes the active south pole of Enceladus. *Science* **311,** 1393–1401 (2006).
2. Spencer, J. R. *et al.* Cassini encounters Enceladus: background and the discovery of a south polar hot spot. *Science* **311,** 1401–1405 (2006).
3. Nimmo, F. & Pappalardo, R. T. Diapir-induced reorientation of Saturn's moon Enceladus. *Nature* **441,** 614–616 (2006).
4. Collins, G. & Goodman, J. C. Internal melting and the shape of Enceladus. *AAS/Division Planet. Sci. Conf.* **38,** abstr. 18.03 (American Astronomical Society, 2006).
5. Kieffer, S. *et al.* A clathrate reservoir hypothesis for Enceladus' south polar plume. *Science* **314,** 1764–1766 (2006).
6. Ingersoll, A. P., Porco, C. C., Helfenstein, P., West, R. A. & the Cassini ISS Team. Models of the Enceladus plumes. *AAS/Division Planet. Sci. Conf.* **38,** abstr. 15.02 (American Astronomical Society, 2006).
7. Nimmo, F. & Gaidos, E. Strike-slip motion and double ridge formation on Europa. *J. Geophys. Res.* **107,** doi:10.1029/2000JE001476 (2002).
8. Prockter, L. M., Nimmo, F. & Pappalardo, R. T. A shear heating origin for ridges on Triton. *Geophys. Res. Lett.* **32,** doi:10.1029/2005GL022832 (2005).
9. Nimmo, F., Spencer, J. R., Pappalardo, R. T. & Mullen, M. E. Shear heating as the origin of the plumes and heat flux on Enceladus. *Nature* doi:10.1038/nature05783 (this issue).
10. Horanyi, M., Burns, J. A. & Hamilton, D. P. The dynamics of Saturn's E ring particles. *Icarus* **97,** 248–259 (1992).
11. Giese, B., Roatsch, T., Wagner, R., Denk, T. & Neukum, G. Topographic models of the icy Saturn satellites. *Eur. Planet. Sci. Congr.* abstr. 119 (2006).
12. Sohl, F., Hussmann, H. & Ziethe, R. Interior structures of Enceladus and Mimas: implications from their densities and equilibrium shapes. *AAS/Division Planet. Sci. Conf.* **38,** abstr. 24.01 (American Astronomical Society, 2006).
13. Hussmann, H., Sohl, F. & Grott, M. Tidal heating in Enceladus And Mimas: implications for their thermal and orbital states. *AAS/Division Planet. Sci. Conf.* **38,** abstr. 24.02 (American Astronomical Society, 2006).
14. Hurford, T. A. *Tides and Tidal Stress: Applications to Europa*. PhD thesis, Univ. Arizona (2005).
15. Melosh, H. J. Global tectonics of a despun planet. *Icarus* **31,** 221–243 (1977).
16. Leith, A. C. & McKinnon, W. B. Is there evidence for polar wander on Europa? *Icarus* **120,** 387–398 (1996).
17. Gammon, P. H., Kiefte, H., Clouter, M. J. & Denner, W. W. Elastic constants of artificial and natural ice samples by Brillouin spectroscopy. *J. Glaciol.* **29,** 433–460 (1983).
18. Hoppa, G., Tufts, B. R., Greenberg, R. & Geissler, P. Strike-slip faults on Europa: global shear patterns driven by tidal stress. *Icarus* **141,** 287–298 (1999).
19. Hoppa, G. V., Greenberg, R., Tufts, B. R. & Geissler, P. E. Plume detection on Europa: locations of favorable tidal stress. *Lunar Planet. Inst. Conf.* **30,** abstr. 1603 (1999).
20. Porco, C. *et al.* Enceladus' jets: particle characteristics, surface source locations, temporal variability, and correlations with thermal hot spots. *Lunar Planet. Inst. Conf.* **38,** abstr. 2310 (2007).
21. Roatsch, T. *et al.* High resolution Enceladus atlas derived from Cassini-ISS images. *Planet. Space Sci.* (submitted).
22. Hurford, T. A., Greenberg, R. & Hoppa, G. V. South polar cycloidal rift on Enceladus. *AAS/Division Planet. Sci. Conf.* **38,** abstr. 18.04 (American Astronomical Society, 2006).
23. Hoppa, G. V., Tufts, B. R., Greenberg, R. & Geissler, P. E. Formation of cycloidal features on Europa. *Science* **285,** 1899–1902 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the NASA Posdoctoral Program (NPP) under the administration of Oak Ridge Associated Universities (ORAU).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.A.H. (hurfordt@core2.gsfc.nasa.gov).

LETTERS

Electronic measurement and control of spin transport in silicon

Ian Appelbaum[1], Biqin Huang[1] & Douwe J. Monsma[2]

The spin lifetime and diffusion length of electrons are transport parameters that define the scale of coherence in spintronic devices and circuits. As these parameters are many orders of magnitude larger in semiconductors than in metals[1,2], semiconductors could be the most suitable for spintronics. So far, spin transport has only been measured in direct-bandgap semiconductors[3–9] or in combination with magnetic semiconductors, excluding a wide range of non-magnetic semiconductors with indirect bandgaps. Most notable in this group is silicon, Si, which (in addition to its market entrenchment in electronics) has long been predicted a superior semiconductor for spintronics with enhanced lifetime and transport length due to low spin–orbit scattering and lattice inversion symmetry[10–12]. Despite this promise, a demonstration of coherent spin transport in Si has remained elusive, because most experiments focused on magnetoresistive devices; these methods fail because of a fundamental impedance mismatch between ferromagnetic metal and semiconductor[13], and measurements are obscured by other magnetoelectronic effects[14]. Here we demonstrate conduction-band spin transport across 10 μm undoped Si in a device that operates by spin-dependent ballistic hot-electron filtering through ferromagnetic thin films for both spin injection and spin detection. As it is not based on magnetoresistance, the hot-electron spin injection and spin detection avoids impedance mismatch issues and prevents interference from parasitic effects. The clean collector current shows independent magnetic and electrical control of spin precession, and thus confirms spin coherent drift in the conduction band of silicon.

Figure 1a illustrates the operating principle and schematic band diagram of our device. Spin injection and detection is based on the attenuation of minority-spin hot electrons in ferromagnetic thin films, as in spin-valve transistors[15,16]. In our device, the spin-valve transistors used for injection and detection each only have a single ferromagnetic base layer, and we define these as 'hot-electron spin transistors'. In step 1, a solid-state tunnel junction injects unpolarized hot electrons from the Al emitter into the ferromagnetic $Co_{84}Fe_{16}$ base, forming emitter current I_e. Spin-dependent hot-electron scattering attenuates minority spin electrons (step 2), so that the electrons transported over the Schottky barrier and into the undoped single-crystal float-zone (FZ)-Si conduction band (forming injected current I_{c1}, the 'first collector current') are polarized, with their spin parallel to the magnetization of the $Co_{84}Fe_{16}$ (step 3)[17]. After vertical transport through the 10-μm-thick undoped Si (step 4), the spin polarization of the conduction-band electrons is detected by a second hot-electron spin transistor. The $Ni_{80}Fe_{20}$ base again uses ballistic hot-electron spin filtering, so the 'second collector current' (I_{c2}, step 5) formed from ballistic transport through the $Ni_{80}Fe_{20}$ and into the n-Si substrate conduction band is dependent on the relative magnetizations of both ferromagnetic layers. When they are parallel, I_{c2} is higher than when they are antiparallel, but only if electron spin polarization is maintained through the undoped Si layer. Therefore, this device is the electron analogue of the photon polarization-analyser experiment in optics.

There are various intrinsic device aspects that allow a clean spin transport signal in I_{c2}, and that make it immune to fringe field-induced magnetoresistance and Hall effects. (1) The exponential spin selective mean free path dependence in the ferromagnetic films create very large spin polarizations. In principle, this can approach 100%,

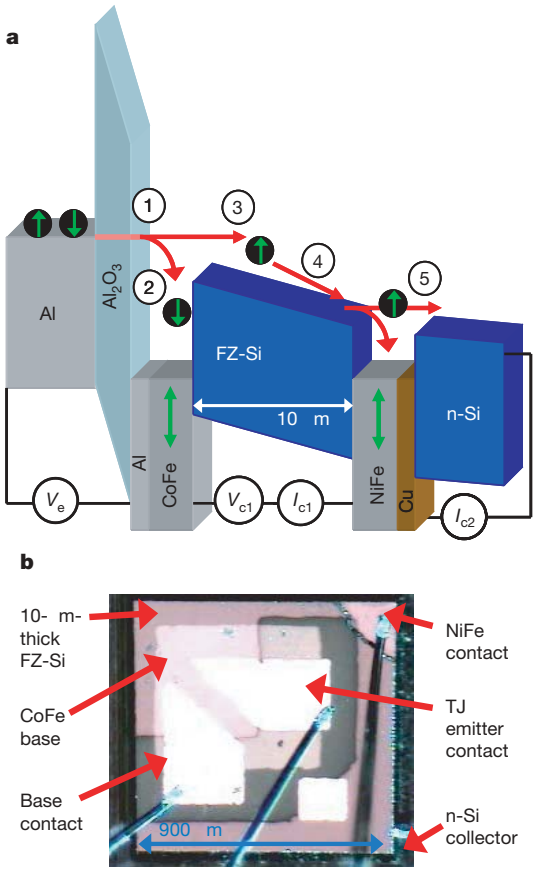


Figure 1 | Illustration of the Si spin transport device. a, Schematic band diagram. At constant emitter voltage (V_e), we measure the 'first collector current' (I_{c1}) at the NiFe contact and 'second collector current' (I_{c2}) at an In contact to the n-Si substrate, under optional voltage bias (V_{c1}) across the single-crystal undoped float-zone Si (FZ-Si) drift region. See text for explanation of sequential transport steps (1)–(5). **b**, A top-down micrograph of a representative wire-bonded Si spin-transport device, showing the device structure, contacts to the spin-injection tunnel junction (TJ) base and emitter, and spin-detector buried NiFe layer.

[1]Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716, USA. [2]Cambridge NanoTech Inc., Cambridge, Massachusetts 02139, USA.

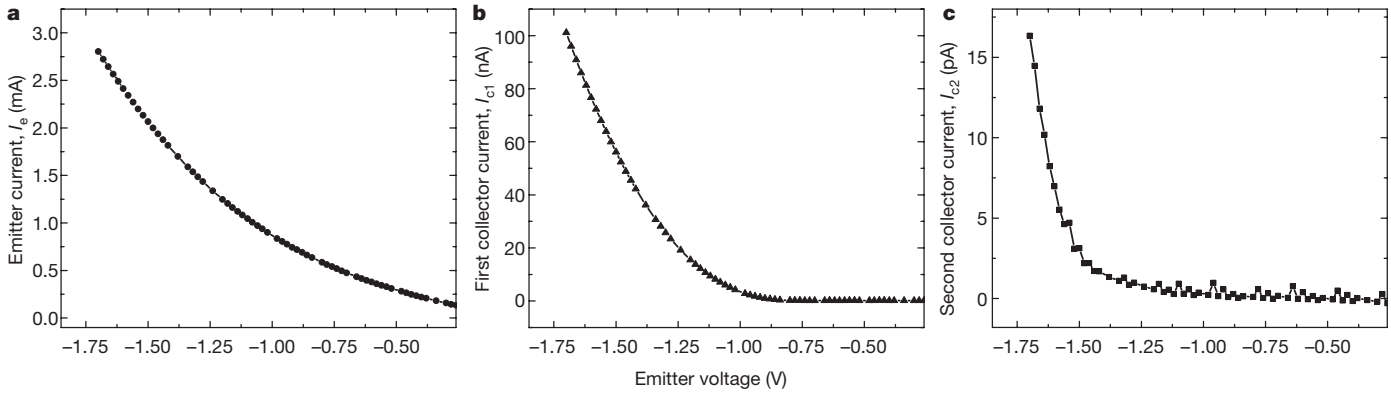


Figure 2 | Simultaneously measured current dependence on tunnel-junction emitter voltage at 85 K. **a**, Emitter current, I_e. **b**, First collector current, I_{c1}, at the $Ni_{80}Fe_{20}$ contact for $V_{c1} = 0$V. **c**, Second collector current, I_{c2}, at an In contact to the n-Si substrate.

allowing effective injection and detection at cryogenic and room temperatures[16]. (2) Because the spin filtering is caused by bulk scattering in the ferromagnetic films, they are easy to reproduce, as there is no interface sensitivity to the spin filtering (as there is, for example, in magnetic tunnel junctions[18]). (3) This device, like a spin-valve transistor, is a high-impedance current source[15,16]. As I_{c2} is driven by I_{c1}, and I_{c1} by I_e, I_{c2} is virtually independent of V_{c1}, the applied voltage across the Si drift region. This also means that any generated Hall voltage in the FZ-Si has no effect on I_{c2}. The underlying background to the insensitivity to resistance and voltage of the FZ-Si is that the potential is screened by the two Schottky barriers on either side, and that the electrons travelling in the FZ-Si conduction band are generated not by an ohmic source, but by hot-electron injection. (4) These devices operate over a wide temperature range, without appreciable change in I_{c2}, despite the fact that the resistivity of undoped Si varies by many orders of magnitude. The insensitivity of I_{c2} to FZ-Si resistance implies that I_{c2} is also insensitive to magnetoresistance in the FZ-Si. In fact, at the temperature we use for measurements here (85 K), the FZ-Si is completely frozen out and its resistivity is $>10^{14}\,\Omega$ cm. This means that there are no thermally or impurity generated electrons, and the only free electrons present are the injected spin polarized electrons.

There are several design aspects that provide our device with a clean spin-transfer current I_{c2}. (1) An undoped FZ-Si device layer is chosen because its extremely low impurity density results in wide Schottky depletion regions and a linear conduction band. This prevents potential wells and long spin dwell times. (2) The device is measured at 85 K to eliminate Schottky leakage currents. (3) Copper (Cu) is used below the $Ni_{80}Fe_{20}$ to provide a low barrier and enable electrons injected by the higher $Si/Ni_{80}Fe_{20}$ Schottky barrier to overcome the Cu/n-Si Schottky barrier. (4) Shape anisotropy of the ferromagnetic films allows us to apply a perpendicular magnetic field for spin precession measurements, without orienting the magnetizations out of plane.

Our fabrication procedures, described in the Methods section, are similar to those of the spin-valve transistor[16], and result in an array of devices like the one wire-bonded and shown in the micrograph displayed in Fig. 1b.

Figure 2a shows the injector tunnel junction current–voltage characteristics, illustrating the expected nonlinear I_e–V_e relationship. Figure 2b shows the simultaneous measurement of injected current I_{c1}–V_e, demonstrating a threshold in I_{c1} at $V_e = -0.8$ V. This represents the 0.8 eV potential energy needed for the electrons to exceed the $Co_{84}Fe_{16}$/FZ-Si Schottky barrier height, and is typical for such metal base transistor-type structures. After vertical transport through 10 μm FZ-Si, some of these electrons travel ballistically through the $Ni_{80}Fe_{20}$/Cu film and into the n-Si collector, resulting in a current (I_{c2}) detected at the In contact to the substrate. This signal rises above our detection limit at an emitter voltage of approximately −1.2 V (Fig. 2c).

In Fig. 3a, in-plane magnetic hysteresis data of I_{c2} at $V_e = -1.8$ V and 85 K are shown. Measurements begin with fully saturated and aligned magnetizations by ramping to our magnetic field maximum. When the magnetic field is swept through zero and changes sign, first the $Ni_{80}Fe_{20}$ switches to align with the field and the magnetizations are antiparallel, resulting in a reduction in I_{c2} of approximately 2%. As the magnetic field passes through the $Co_{84}Fe_{16}$ switching field, the magnetizations again align and the higher collector current is

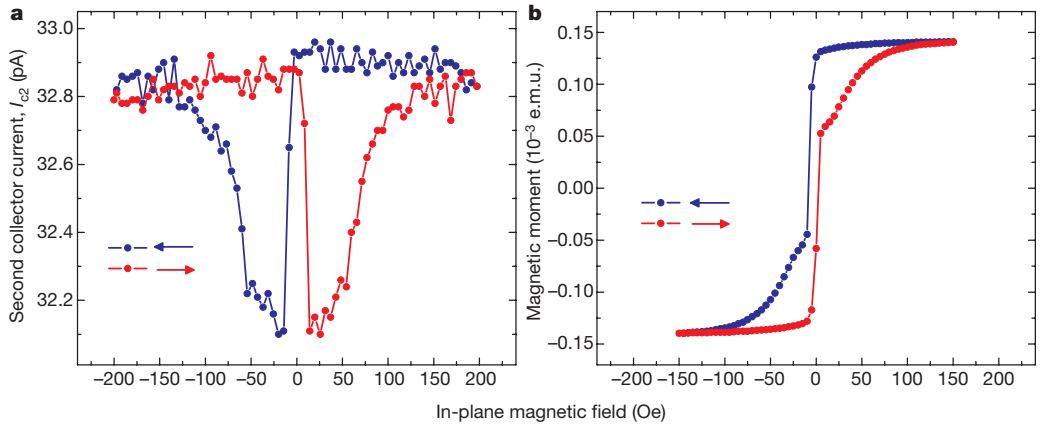


Figure 3 | In-plane magnetic field dependence at 85K. **a**, Second collector current, I_{c2}, at constant emitter bias $V_e = -1.8$ V and $V_{c1} = 0$ V, showing a clear ~2% spin-valve effect. **b**, SQUID magnetometer measurements, showing switching fields consistent with the behaviour seen in **a**. Arrows indicate magnetic field sweep direction.

regained by approximately 125 Oe, consistent with the magnetic hysteresis data shown in Fig. 3b. The symmetric magnetic-field dependence of I_{c2} upon reversal of sweep direction in Fig. 3a indicates that the electron spin maintains polarization while travelling through 10 μm of FZ-Si. Results similar to Fig. 3a are also found when the emitter tunnel junction is operated in constant current mode, and many devices have consistently been measured with substantially similar results.

In Fig. 4a, we show the dependence of I_{c2} on magnetic field perpendicular to the film plane with $V_{c1} = 0$ V. The measurement begins with the field at −5,000 Oe, where a small in-plane component is sufficient to saturate the in-plane $Ni_{80}Fe_{20}$ and $Co_{84}Fe_{16}$ magnetizations in a parallel configuration. Following the red line to the right towards smaller field values, we notice a peak near −700 Oe and a dip near −350 Oe. Once the magnetic field increases towards positive values past zero, the small in-plane component of external magnetic field switches the $Ni_{80}Fe_{20}$ magnetization, causing an antiparallel magnetization configuration and inverting the magnitudes of the oscillation. Reversing the field scan direction (blue line) results in similar features reflected on the magnetic field axis, but with a deeper dip at small negative field due to sharper $Ni_{80}Fe_{20}$ switching in that direction.

Although $V_{c1} = 0$ V in Fig. 4a, a vertical drift field (due to a voltage drop caused by emitter current flowing through the $Al/Co_{84}Fe_{16}$ thin base film) exists in the FZ-Si. Under these conditions, where drift is the primary transport mechanism, the magnetic fields of the extrema seen in Fig. 4a can be identified approximately as the conditions for π and 2π precession. The precessing spin direction, projected on the measurement axis defined by the $Ni_{80}Fe_{20}$ magnetization, causes oscillation in I_{c2}.

Because of the action of random diffusion, spin dephasing occurs with higher precession angles, and results in damped higher-order oscillations (Hanle effect)[19]. Therefore, just two multiples of π precession angle are seen. A useful advantage of the four-terminal design

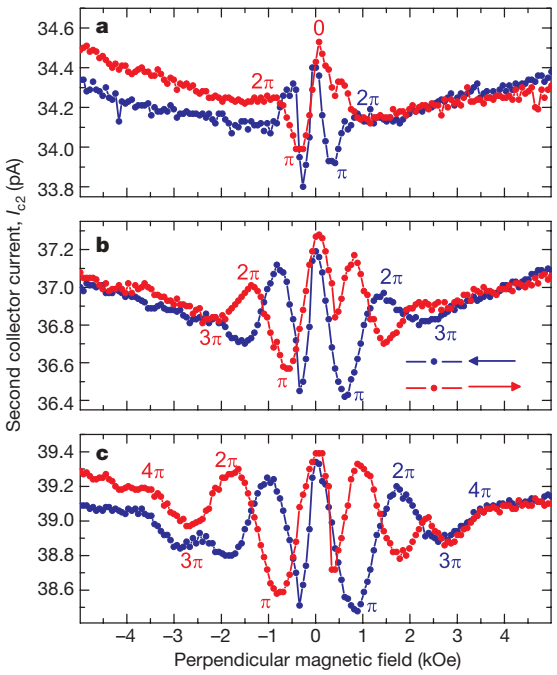


Figure 4 | Spin precession and dephasing in a perpendicular magnetic field at constant emitter voltage $V_e = -1.8$ V and 85 K. a, Second collector current, I_{c2}, in zero applied voltage V_{c1}. **b**, **c**, As **a** but under an accelerating voltage bias of 0.5 V (**b**) and of 1.0 V (**c**). At higher accelerating voltages, the spin polarized electron transport time is reduced in the increased drift field, so the precession angle at fixed magnetic field is smaller, causing an increased precession period and revealing the presence of precession angles up to 4π.

and rectifying Schottky barriers on either side of the FZ-Si in our device is that I_{c2} is virtually independent of voltage across the drift region. Therefore, the drift electric field can be tuned to change the spin polarized electron transit time between injection and detection with applied voltage V_{c1}. When the electric drift field is increased with an applied voltage bias, the transit time is reduced, and the precession angle at any fixed magnetic field is consequently also reduced. This pushes the extrema to higher values of perpendicular magnetic fields, clearly shown in Fig. 4b and c at respectively 0.5 V and 1.0 V accelerating voltage. Under these conditions of higher accelerating electric field, drift is even more dominant than in Fig. 4a, and precession angles up to 4π can be seen. The precession extrema for parallel magnetization configurations are labelled with their precession angle in Fig. 4a–c. In this way, maxima correspond to even multiples of π and minima correspond to odd multiples of π. The large number of precession extrema and their expected electric field dependence are conclusive evidence of strong coherent spin transport in the FZ-Si.

Using the spin precession frequency $g\mu_B/h = 2.8\,\mathrm{MHz\,Oe^{-1}}$ (where g is the electron spin g-factor, μ_B is the Bohr magneton, and h Planck's constant; ref. 20), the transit time can be calculated from the magnetic field value of the first extremum, where the average precession angle is approximately π rad. Using ±350 Oe as the position of the extrema in Fig. 4a, we have $\tau_{transit} = \pi/(2\pi \times 2.8 \times 10^6 \times 350) \approx 0.5$ ns, which gives an approximately $2 \times 10^6\,\mathrm{cm\,s^{-1}}$ drift velocity through the 10-μm-thick FZ-Si[21]. The positions of the precession extrema in Fig. 4b and c indicate transit times of ~0.3 ns and ~0.2 ns, respectively.

As the electron transit time e is controlled by applied voltage V_{c1}, the normalized magnetocurrent change in I_{c2} (ΔI_{c2}) can be used to deduce the spin lifetime in FZ-Si. Assuming a simple exponential decay law, and using the measured transit times from Fig. 4a–c, we can calculate the spin lifetime τ_{sf} (at 85 K) using ratios of

$$\frac{\Delta I_{c2}(V_{c1})}{I_{c1}(V_{c1})} \propto \exp\left[-\frac{\tau_{transit}(V_{c1})}{\tau_{sf}}\right]$$

for each pair of data. With the maximum fluctuations shown in Fig. 4a–c, and using simultaneously-measured $I_{c1}(0\mathrm{V}) = 118$ nA, $I_{c1}(0.5\mathrm{V}) = 132$ nA and $I_{c1}(1.0\mathrm{V}) = 139$ nA, we calculate $\tau_{sf} \approx 1$ ns. This value is consistent with spin lifetimes in the direct-bandgap semiconductor GaAs measured at similar temperatures using optical techniques[3]. However, we must consider this value a lower bound, with the actual lifetime potentially much longer:[22] the applied voltage V_{c1} could alter the magnetocurrent magnitude ΔI_{c2} separately from the finite spin lifetime and variable transit time effect, artificially reducing our estimate.

It is possible to use alternative spin injection schemes with our detection method, such as direct tunnel injection[23–25]. However, in that case the injection current is not decoupled from drift field and thus independent control of spin precession by magnetic field and electric drift field is compromised; on the other hand, the total signal will be larger. Finally, we hope that our hot-electron injection/detection scheme will be implemented in lateral geometries, to advance the development of future spin-based integrated circuits.

METHODS

Room-temperature ultrahigh-vacuum wafer bonding[16] is used to assemble the semiconductor–metal–semiconductor spin detection structure. First, a 4-nm-thick Cu film is deposited via thermal evaporation on a 1 Ω cm 50-mm-diameter n-Si wafer to achieve a low second collector Schottky barrier. Then, during deposition of 2 nm of $Ni_{80}Fe_{20}$ on both this wafer and on a 50-mm-diameter silicon-on-insulator (SOI) wafer with a 10 μm undoped single-crystal FZ-Si device layer, the wafers are pressed together *in situ*, forming a cohesive bond and a single 4-nm-thick $Ni_{80}Fe_{20}$ film. Transmission electron microscopy analysis of similarly bonded interfaces is shown in ref. 16. After SOI substrate removal and mesa patterning using the 2-μm-thick buried oxide as a mask, a tunnel junction structure is deposited using electron beam deposition of a 5 nm $Co_{84}Fe_{16}$ and 5 nm Al bilayer base (500 μm × 500 μm) and (after ultraviolet

ozone oxidation to form the tunnel barrier) two Al contacts. For isolation, a wafer saw is used to cut trenches partially through the n-Si collector substrate, forming an array of devices each 900 μm × 900 μm.

Pinholes in these deposited layers cannot result in spurious measurements of I_{c2}: (1) if they exist in the tunnel oxide, the emitter tunnel junction is shorted and the device is clearly inoperable; (2) if they exist in the Al/$Co_{84}Fe_{16}$ layer, the emitter forms a rectifying Schottky barrier there and the device is unaffected; and (3) pinholes in the single-crystal FZ-Si layer are not possible.

As fabricated, the two Al contacts both form tunnel junctions with the Al/$Co_{84}Fe_{16}$ base. Ramping a current source through them in series breaks one and leaves the other intact. Contact for I_{c1} is to the $Ni_{20}Fe_{80}$ buried layer, and contact for I_{c2} is with cold-pressed indium to the n-Si substrate. This current has been measured with a Keithley 236 source-measure unit, without any other additions or special efforts.

Received 22 December 2006; accepted 30 March 2007.

1. Jedema, F. J., Heersche, H. B., Filip, A. T., Baselmans, J. J. A. & van Wees, B. J. Electrical detection of spin precession in a metallic mesoscopic spin valve. *Nature* **416,** 713–717 (2002).
2. Valenzuela, S. O. & Tinkham, M. Direct electronic measurement of the spin Hall effect. *Nature* **442,** 176–179 (2006).
3. Kikkawa, J. M. & Awschalom, D. D. Resonant spin amplification in n-type GaAs. *Phys. Rev. Lett.* **80,** 4313–4316 (1998).
4. Kikkawa, J. M. & Awschalom, D. D. Lateral drag of spin coherence in gallium arsenide. *Nature* **397,** 139–141 (1999).
5. Kato, Y., Myers, R. C., Gossard, A. C. & Awschalom, D. D. Coherent spin manipulation without magnetic fields in strained semiconductors. *Nature* **427,** 50–53 (2003).
6. Stephens, J. *et al.* Spin accumulation in forward-biased MnAs/GaAs Schottky diodes. *Phys. Rev. Lett.* **93,** 097602 (2004).
7. Crooker, S. A. *et al.* Imaging spin transport in lateral ferromagnet/ semiconductor structures. *Science* **309,** 2191–2195 (2005).
8. Lou, X. *et al.* Electrical detection of spin accumulation at a ferromagnet-semiconductor interface. *Phys. Rev. Lett.* **96,** 176603 (2006).
9. Lou, X. *et al.* Electrical detection of spin transport in lateral ferromagnet–semiconductor devices. *Nature Phys.* **3,** 197–202 (2007).
10. Zutic, I., Fabian, J. & Erwin, S. C. Spin injection and detection in Si. *Phys. Rev. Lett.* **97,** 026602 (2006).
11. Zutic, I., Fabian, J. & Das Sarma, S. Spintronics: Fundamentals and applications. *Rev. Mod. Phys.* **76,** 323–410 (2004).
12. Tyryshkin, I. M., Lyon, S. A., Astashkin, A. V. & Raitsimring, A. M. Electron spin relaxation times of phosphorus donors in Si. *Phys. Rev. B* **68,** 193207 (2003).
13. Schmidt, G., Ferrand, D., Molenkamp, L. W., Filip, A. T. & van Wees, B. J. Fundamental obstacle for electrical spin injection from a ferromagnetic metal into a diffusive semiconductor. *Phys. Rev. B* **62,** R4790–R4793 (2000).
14. Monzon, F. G., Tang, H. X. & Roukes, M. L. Magnetoelectronic phenomena at a ferromagnet-semiconductor interface. *Phys. Rev. Lett.* **84,** 5022–5022 (2000).
15. Monsma, D. J., Lodder, J. C., Popma, ThJA & Dieny, B. Perpendicular hot electron spin-valve effect in a new magnetic field sensor: The spin valve transistor. *Phys. Rev. Lett.* **74,** 5260–5263 (1995).
16. Monsma, D. J., Vlutters, R. & Lodder, J. C. Room temperature-operating spin-valve transistors formed by vacuum bonding. *Science* **281,** 407–409 (1998).
17. Jiang, X. *et al.* Optical detection of hot-electron spin injection into GaAs from a magnetic tunnel transistor source. *Phys. Rev. Lett.* **90,** 256603 (2003).
18. Zega, T. J. *et al.* Determination of interface atomic structure and its impact on spin transport using Z-contrast microscopy and density-functional theory. *Phys. Rev. Lett.* **96,** 196101 (2006).
19. Johnson, M. & Silsbee, R. H. Interfacial charge-spin coupling: Injection and detection of spin magnetization in metals. *Phys. Rev. Lett.* **55,** 1790–1793 (1985).
20. Portis, A. M., Kip, A. F., Kittel, C. & Brattain, W. H. Electron spin resonance in a silicon semiconductor. *Phys. Rev.* **90,** 988–989 (1953).
21. Jacoboni, C., Canali, C., Ottaviani, G. & Quaranta, A. A. A review of some charge transport properties of Si. *Solid State Electron.* **20,** 77–89 (1977).
22. Tyryshkin, A. M., Lyon, S. A., Jantsch, W. & Schäffler, F. Spin manipulation of free two-dimensional electrons in Si/SiGe quantum wells. *Phys. Rev. Lett.* **94,** 126802 (2005).
23. Jiang, X. *et al.* Highly spin polarized room temperature tunnel injector for semiconductor spintronics using MgO (100). *Phys. Rev. Lett.* **94,** 056601 (2005).
24. Smith, D. L. & Silver, R. N. Electrical spin injection into semiconductors. *Phys. Rev. B* **64,** 045323 (2002).
25. Hanbicki, A. T., Jonker, B. T., Itskos, G., Kioseoglou, G. & Petrou, A. Efficient electrical spin injection from a magnetic metal/tunnel barrier contact into a semiconductor. *Appl. Phys. Lett.* **80,** 1240–1242 (2002).

Acknowledgements We acknowledge assistance during fabrication from I. Altfeder, SQUID measurements by G. Hadjipanayis and A. Gabay, and use of the wafer saw from K. Goossen. This work is supported by ONR and DARPA/MTO.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.A. (appelbaum@ee.udel.edu).

LETTERS

Annealing-induced interfacial toughening using a molecular nanolayer

Darshan D. Gandhi[1], Michael Lane[3], Yu Zhou[2], Amit P. Singh[1], Saroj Nayak[2], Ulrike Tisch[4], Moshe Eizenberg[4] & Ganapathiraman Ramanath[1]

Self-assembled molecular nanolayers (MNLs) composed of short organic chains and terminated with desired functional groups are attractive for modifying surface properties for a variety of applications. For example, organosilane MNLs are used as lubricants[1], in nanolithography[2], for corrosion protection[3] and in the crystallization of biominerals[4]. Recent work has explored uses of MNLs at thin-film interfaces, both as active components in molecular devices[5], and as passive layers, inhibiting interfacial diffusion[6–8], promoting adhesion[9,10] and toughening brittle nanoporous structures[11]. The relatively low stability of MNLs on surfaces at temperatures above 350–400 °C (refs 12, 13), as a result of desorption[14] or degradation, limits the use of surface MNLs in high-temperature applications. Here we harness MNLs at thin-film interfaces at temperatures higher than the MNL desorption temperature to fortify copper–dielectric interfaces relevant to wiring in micro- and nano-electronic devices. Annealing Cu/MNL/SiO_2 structures at 400–700 °C results in interfaces that are five times tougher than pristine Cu/SiO_2 structures, yielding values exceeding ~20 J m^{-2}. Previously, similarly high toughness values have only been obtained using micrometre-thick interfacial layers[15–17]. Electron spectroscopy of fracture surfaces and density functional theory modelling of molecular stretching and fracture show that toughening arises from thermally activated interfacial siloxane bridging that enables the MNL to be strongly linked to both the adjacent layers at the interface, and suppresses MNL desorption. We anticipate that our findings will open up opportunities for molecular-level tailoring of a variety of interfacial properties, at processing temperatures higher than previously envisaged, for applications where microlayers are not a viable option—such as in nanodevices or in thermally resistant molecular-inorganic hybrid devices.

Molecular nanolayers of 3-mercapto-propyl-tri-methoxy-silane (MPTMS) were adsorbed on an oxidized silicon substrate by immersion in an MPTMS solution (see Methods). Variable angle spectroscopic ellipsometry and variable take-off angle X-ray photoelectron spectroscopy (XPS) measurements showed that MPTMS MNLs are 0.9 ± 0.1 nm thick, with the thiol termini pointing outwards. This thickness is in good agreement with the theoretical length of MPTMS, confirming the formation of a monolayer. Four-point-bending fracture tests[10,18] were carried out on thin-film sandwiches of Si/epoxy/Ta/Cu/MPTMS/SiO_2/Si structures (Fig. 1) to determine the interface toughness, G (which is also known as the interface debonding energy or the strain energy release rate). Figure 1 also shows the measured interface toughness for such structures, with and without an MPTMS layer, after annealing between 100 and 700 °C. (See Supplementary Fig. 1 for the load–displacement plots from which the data were obtained.)

As-prepared thin-film structures with pristine Cu/SiO_2 interfaces—that is, without MPTMS MNLs—exhibit an average interface toughness of $G_{reference} = 3.8 \pm 1$ J m^{-2}. There is no observable change in toughness upon annealing. MPTMS-treated interfaces at room temperature show interface toughness of up to a factor of 3 higher, $G_{MPTMS} = 9.1 \pm 1$ J m^{-2}, in agreement with our prior work[10]. Annealing the MPTMS-treated samples between 100 °C $\leq T_{anneal} <$ 400 °C decreases G_{MPTMS} to 3.5 ± 1 J m^{-2}. For $T_{anneal} \geq 400$ °C, however, we observe a monotonic increase in G_{MPTMS}. At 500 °C, $G_{MPTMS} = 20 \pm 2$ J m^{-2}, which is more than five times larger than

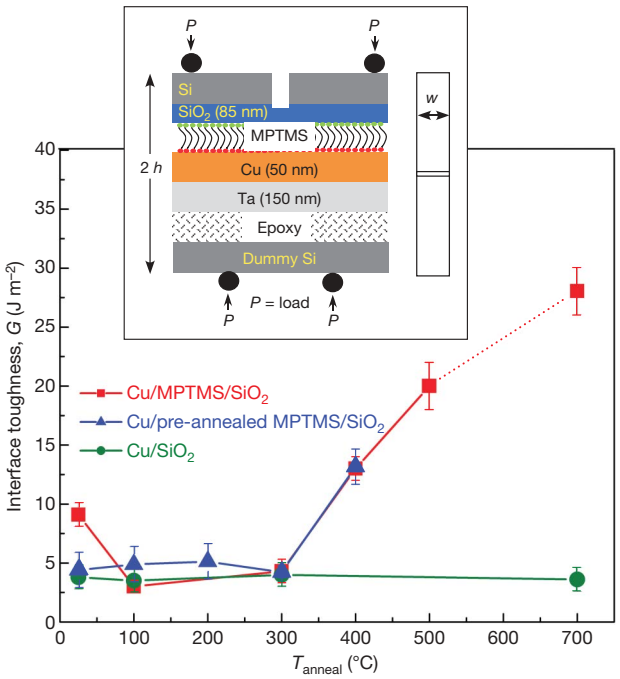


Figure 1 | Effect of thermal annealing on the interface toughness of Cu/MPTMS/SiO_2 and Cu/SiO_2 structures. Inset, diagram of the four-point-bend test structure; for brevity, this legend and the figure key omit mention of the outer Si/epoxy/Ta/...../Si parts of the test structure. The sample thickness and width are given by $2h$ and W, respectively, and P is the load applied during four point bending tests. The graph shows average interface toughness, G, plotted as a function of annealing temperature, T_{anneal}, for Cu/MPTMS/SiO_2 and Cu/SiO_2 structures. The dotted line for Cu/MPTMS/SiO_2 structures for $T_{anneal} > 500$ °C represents a lower limit of the interface toughness. The blue data points are from samples where MPTMS was pre-annealed before Cu deposition. Data points, mean of results from six different samples; error bars, ±s.d.

[1]Materials Science & Engineering Department, [2]Physics Department, Rensselaer Polytechnic Institute, Troy, New York 12180, USA. [3]IBM Microelectronics, T.J. Watson Research Center, Yorktown Heights, New York 10598, USA. [4]Department of Materials Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel.

that measured for Cu/SiO_2 interfaces without an MNL. Such large interfacial toughening through the use of a nanometre-thick molecular layer is unprecedented. This result is particularly remarkable because toughening is achieved by annealing the samples with MPTMS at temperatures above which MPTMS is known to thermally desorb, for example, 350–400 °C (ref. 14). Our approach of using nanolayers at interfaces combined with thermal annealing therefore offers an alternative to micrometre-thick metallic[19] or polymeric[16] layers currently used to obtain similar toughness values[20,21]. We demonstrate below that the bonding between the thiol termini of the MPTMS MNL and the Cu overlayer inhibits molecular desorption as expected, and allows toughening to occur by annealing-induced strengthening of the MNL/SiO_2 interface via siloxane bridging.

For $T_{anneal} \leq 500$ °C, delamination occurs at the molecularly tailored Cu/MPTMS/SiO_2 interface, as revealed by XPS spectra (Fig. 2a) from the two fracture surfaces, showing spectral signatures of Cu (from the Cu fracture surface), and silicon and oxygen (from silica), respectively. The S 2*p* signal from the nanometre-thick MNL of MPTMS is below the detection limit in survey spectra in Fig. 2, but is discernible in the multiplex scans shown in Fig. 3. The interface toughness measured from samples with MPTMS MNLs subjected to a pre-annealing treatment at 100 °C before Cu deposition follows a similar trend, but with a key difference. No toughening is observed in the Cu/MPTMS/SiO_2 structure with the pre-annealed MPTMS, in contrast to the threefold toughness increase observed in such samples without the pre-anneal. The reason for this difference is discussed later.

For $T_{anneal} > 500$ °C, the Si/epoxy/Ta/Cu/MPTMS/SiO_2/Si stack fractures cohesively inside the epoxy, instead of at the molecularly tailored Cu/MPTMS/SiO_2 interface. This result is verified by XPS spectra from the two fracture surfaces (Fig. 2b) consistently showing identical core-level peaks, indistinguishable from that obtained from the epoxy. Neither fracture surface shows any Cu 2*p* or S 2*p* states. Cohesive failure in the epoxy for $T_{anneal} > 500$ °C suggests that the Cu/MPTMS/SiO_2 interface may not be the weakest one. It is at present unknown if fracture in the epoxy corresponds to the energetically most favourable pathway, or is determined by the dynamics of crack propagation. If we assume that epoxy fracture occurs by the energetically most favoured path in structures annealed to 700 °C, the Cu/MPTMS/SiO_2 interface toughness is at least 28 J m^{-2} (measured for epoxy failure), corresponding to sevenfold toughening. In this scenario, the interface toughness measured for $T_{anneal} > 500$ °C provides only a lower limit of the Cu/MPTMS/SiO_2, and hence is denoted by a dotted line in Fig. 1.

Detailed analysis of core-level spectra from fracture surfaces of samples subjected to $T_{anneal} \leq 500$ °C reveals that fracture occurs at the MNL/SiO_2 interface. Figure 3a and b shows the S 2*p* and Si 2*p* core-level bands from the Cu and silica fracture surfaces, respectively. The S 2*p* core-level signature from the thiol in the MPTMS molecule is seen only on the Cu fracture surface, and is not observed on the silica fracture surface. The S 2*p* state at 161.2 eV confirms Cu–S bonding between the MNL and the Cu overlayer[22]. On the Cu fracture surface, the Si 2*p* band centred at 102.8 eV corresponds to the methoxysilanes in MPTMS[23], unlike the silica fracture surface where the Si 2*p* band is centred at 103.4 eV. The Si 2*p* spectral signature from the Cu fracture surface is identical to that seen in outward-pointing methoxysilanes in MPTMS MNLs bound to a Au surface by thioligation (Fig. 3b). These observations indicate that delamination occurs at the MPTMS/SiO_2 interface for $T_{anneal} \leq 500$ °C. Thioligation of Cu with MPTMS achieved in the as-prepared samples[10] prevents molecular desorption, but cannot explain the large annealing-induced interface toughening because core-level Cu 2*p* and S 2*p* bands are identical for different T_{anneal} (Fig. 3c). Moreover, annealing MPTMS/SiO_2 structures before, or after, Cu deposition (Fig. 1) results in identical changes in the toughness behaviour of Cu/MPTMS/SiO_2 structures.

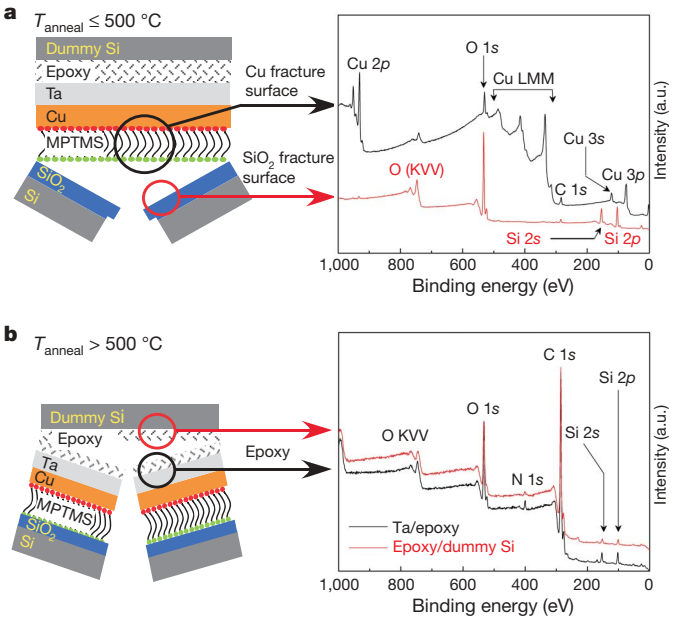


Figure 2 | Core-level spectra from fracture surfaces. Survey XPS scans from samples annealed at **a**, ≤500 °C and **b**, >500 °C. Silica fracture surfaces from samples with $T_{anneal} \leq 500$ °C show Si, O and adventitious C along with O KVV Auger peaks, whereas the Cu fracture surfaces show Cu, O and C core-level bands along with O KVV and Cu LMM signatures. Both the fracture surfaces of all samples with $T_{anneal} > 500$ °C show identical O, N, C and Si core-level bands and O KVV Auger peaks, corresponding to the epoxy, indicating cohesive fracture inside the epoxy.

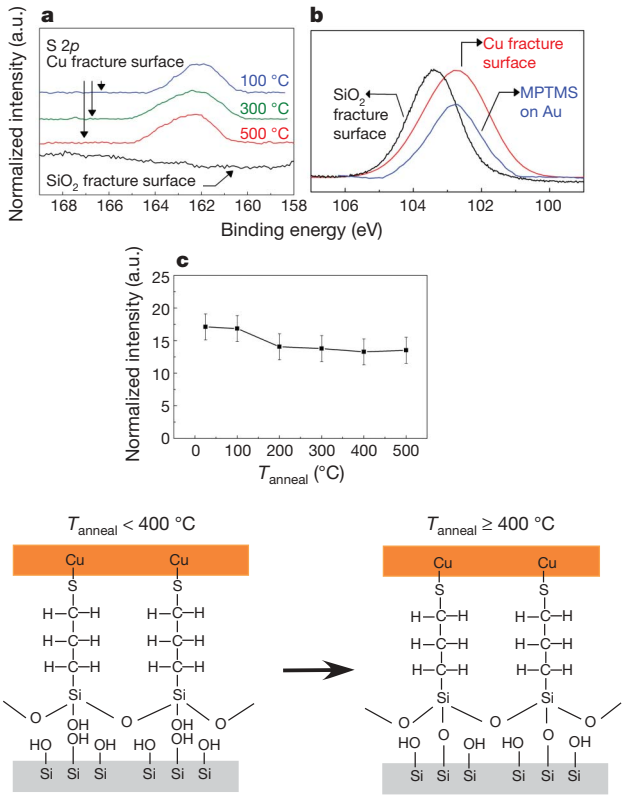


Figure 3 | High-resolution core-level spectra from fracture surfaces, and a schematic illustration of the primary toughening mechanism. **a**, S 2*p* bands from Cu fracture surfaces of samples for $T_{anneal} \leq 500$ °C, shown with a reference spectrum from the silica fracture surface. **b**, Si 2*p* bands from Cu and silica fracture surfaces, respectively, shown with a reference spectrum from an MPTMS MNL on Au. **c**, S 2*p* intensity, normalized to the Cu 2*p* peak, plotted as a function of T_{anneal}. Data points, mean from four data sets; error bars, ±s.d. **d**, Schematic illustration of the temperature-assisted interfacial toughening mechanism. See text for details.

This result indicates that annealing-induced changes in the interfacial mechanical properties are due to chemical changes at the MPTMS/SiO_2 interface.

We attribute annealing-induced interface weakening for $T_{anneal} < 400\,^{\circ}C$, and large interface toughening for $T_{anneal} \geq 400\,^{\circ}C$, to siloxane (Si–O–Si) bond disruption, and stabilization, respectively, at the MPTMS/SiO_2 interface. Prior work[24] on silica glasses has shown that siloxane bonds formed by the dehydration of silanol (Si–OH) groups below 400 °C are strained and susceptible to rehydration, and revert to silanols upon cooling. We expect such reversion of strained siloxane bridges into silanols at the MNL/SiO_2 interface to decrease the Cu/MPTMS/SiO_2 interface toughness, as we indeed observe for $100\,^{\circ}C \leq T_{anneal} < 400\,^{\circ}C$. Siloxane bond breakage does not occur in MPTMS MNLs that are not subjected to any annealing treatment, presumably because of the low thermal activation for siloxane bond breakage by hydration at room temperature. For $T_{anneal} \geq 400\,^{\circ}C$, the scenario is completely different because of the onset of irreversible dehydration, that is, formation of strain-relaxed siloxane bridges that do not revert to silanol groups upon cooling[24]. The resultant increase in the number of strain-free siloxane bridges provides the primary mechanism for interface toughening, as explained later. The monotonic toughness increase with increasing T_{anneal} is probably due to the increase in the number of siloxane bridges formed via dehydration, that is, the increase in the extent of irreversible dehydration with annealing temperature. Figure 3d depicts key features of the interfacial reactions responsible for annealing-induced interfacial weakening below 400 °C, and toughening above 400 °C. Whereas Cu–S bonding occurs at room temperature, stable siloxane bridges are formed at the methoxysilane/SiO_2 interface only upon annealing above ~400 °C. Thus, we attribute the observed temperature-induced toughening to the formation of strong chemical bonds at both Cu/MNL and MNL/SiO_2 interfaces.

The energetics of stretching and fracture of an individual MPTMS molecule, determined by first-principles density functional theory calculations, support our attribution of interface toughening to siloxane bridging at the MPTMS/SiO_2 interface. In these calcualtions, we constrained an MPTMS molecule between a Cu overlayer and a silica underlayer (Fig. 4 inset). The Cu was thioligated, and the methoxy terminal groups in MPTMS were bound to silica via different numbers of covalent siloxane bridges (i) and hydrogen bonds (j), such that $0 \leq i \leq 3$ and $j = 3 - i$. The system energy was calculated as a function of Cu–SiO_2 separation after optimizing the molecule to the ground state for each interface chemistry configuration (see Methods for details).

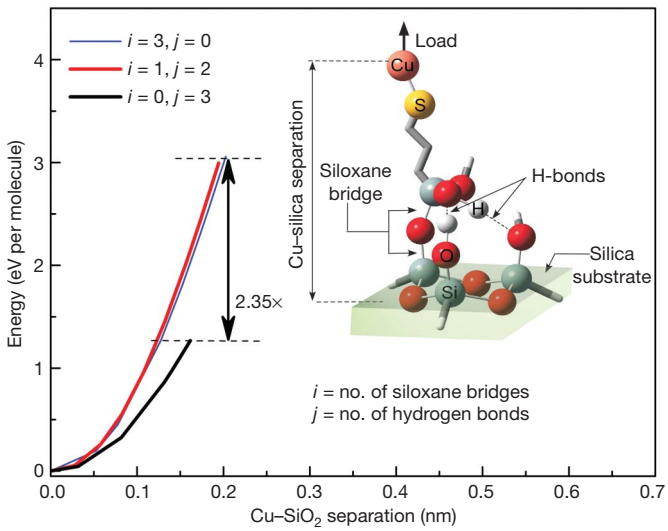


Figure 4 | Calculated system energy as a function of molecular stretching for different combinations of siloxane and hydrogen bonds at the MPTMS/SiO_2 interface. System energy was determined by first-principles density functional theory calculations. Inset, schematic depiction of an MPTMS molecule bonded to the silica underlayer by one siloxane bridge ($i = 1$, 33% irreversible dehydration) and two hydrogen bonds ($j = 2$). The arrow indicates the direction of stretching. The graph shows system energy plotted as a function of Cu overlayer/SiO_2 underlayer separation for three different interface bonding chemistries; each curve is plotted up to the fracture point.

Figure 4 plots the calculated system energy versus Cu–SiO_2 separation for three representative bonding configurations at the MPTMS/SiO_2 interface, namely, $i = 0$, 1 and 3 (and $j = 3 - i$). For MPTMS molecules anchored to silica by hydrogen-bonding ($i = 0$, no siloxane bonds, 0% irreversible dehydration), fracture occurs at the MPTMS/SiO_2 interface. This result is consistent with XPS fracture surface analysis of samples showing lowest interface toughness ($G_{MPTMS} \approx 3\,J\,m^{-2}$, for example) for $100\,^{\circ}C \leq T_{anneal} < 400\,^{\circ}C$. Introducing even one siloxane bond ($i > 1$, 33% irreversible dehydration) shifts the point of bond fracture to the Cu/MPTMS interface and increases the fracture energy by a factor of more than 2.35, underscoring the central role of siloxane bridging in interface toughening. These results, therefore, clearly support our attribution of the experimentally observed toughening to irreversible siloxane bridging for $T_{anneal} \geq 400\,^{\circ}C$. Replacing the other two hydrogen bonds (equivalent to 67% and 100% irreversible dehydration) by siloxane bridges in our theoretical calculations does not result in further toughening, suggesting that one siloxane bridge (33% irreversible hydration) between each MPTMS molecule and the silica underlayer is sufficient to provide toughening. Thus, our results show that irreversible siloxane bridging is an important toughening mechanism, and provide an estimate of the degree of irreversible bridging for a single molecule.

The experimentally measured fracture toughness consists of a number of different energy-absorbing processes. The most fundamental of these is the work of adhesion, which relates directly to the chemical bonding at the Cu/MNL/SiO_2 interface. Interface bond strengthening not only enhances the work of adhesion, but also is a necessary condition for the activation of secondary energy-absorption processes—for example, molecular stretching, lateral molecular interactions[25], and plastic energy dissipation in adjacent films[21,26]. Because our results show that Cu–S bonds at the interface are insensitive to thermal annealing, we propose that annealing-induced siloxane bridging is the primary toughening mechanism. The grain size and texture of the annealed and as-prepared Cu films are identical within experimental uncertainty (Supplementary Fig. 4), ruling out the possibility of toughening due to increased plastic deformation resulting from annealing-induced grain growth. However, siloxane-bridging-induced adhesion increase need not necessarily be the largest contributor to the fracture toughness, when compared to the secondary energy-absorption processes enabled by the increased interfacial strength[26]. We further note that the theoretical prediction of failure at the Cu/MPTMS interface upon siloxane bridging is contrary to our experimentally observed failure at the MPTMS/SiO_2 interface for $T \leq 500\,^{\circ}C$, despite the bond energies of C–C (6.3 eV) and Cu–S (2.86 eV) being significantly lower than that of Si–O (8.3 eV). This apparent discrepancy is due to our theoretical calculations not accounting for the influence of environmental factors such as pH and moisture, which are expected to prematurely break siloxane bonds (for example, by stress corrosion cracking[27]) in our experiments.

Our approach—using a nanolayer interfacial toughener—offers an opportunity to obviate the thicker adhesion promoters currently used, and opens up new possibilities for molecular level tailoring of thin-film interface stability and properties by exploiting high-temperature interactions and integrating MNLs into micro- and nano-device technologies.

METHODS SUMMARY

MPTMS MNLs were adsorbed on n-type Si(001) wafers with an 85-nm-thick thermal oxide cap layer, as described elsewhere[10]. The MNL thickness was measured by ellipsometry, and verified by variable take-off angle XPS

(Supplementary Fig. 2). After MNL formation, one set of samples was used as-is for further processing, while a second set was annealed in air for 10 min at 100 °C. XPS showed no observable oxidation of the thiol termini upon this pre-annealing treatment. Layers of 50-nm-thick Cu and 150-nm-thick Ta were sputter-deposited successively on both sets of MNL samples in a d.c. magnetron tool, without vacuum break, and annealed at $100\ ^{\circ}\mathrm{C} \leq T_{\mathrm{anneal}} \leq 700\ ^{\circ}\mathrm{C}$ for 30 min in 5×10^{-7} torr vacuum. Unannealed samples were also investigated for comparison. Four-point bend tests were carried out on Si(001)/epoxy/Ta/Cu/MPTMS/SiO_2/Si(001) test structures at a 0.01 μm s^{-1} strain rate in a high-stiffness micromechanical system equipped with a piezoelectric actuator. The test structures were prepared as described in detail elsewhere[10]. The interface toughness, G, was determined using the method of ref. 18. The interface chemistry of the fracture surfaces was examined by XPS; instrument details are provided in Supplementary Fig. 2. Theoretical simulations of molecular stretching and failure were carried out using a density functional theory method based on the B3LYP gradient corrected functional, which is well-suited to accurately capture long-range interactions such as hydrogen bonding[28]. We used 6-31+G* basis sets and the Gaussian 03 code. Our calculations consider only a single molecule, and do not capture intermolecular interactions or the effect of moisture and other environmental factors. The location of fracture was determined from an analysis of the changes in rate and sign of C–C, Si–O and Cu–S bond lengths with respect to Cu–SiO_2 separation (Supplementary Fig. 3).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 August 2006; accepted 30 March 2007.

1. Qian, L., Tian, F. & Xiao, X. Tribological properties of self-assembled monolayers and their substrates under various humid environments. *Tribol. Lett.* **15,** 169–176 (2003).
2. Yang, X. M., Peters, R. D., Kim, T. K. & Nealey, P. F. Proximity X-ray lithography using self-assembled alkylsiloxane films: Resolution and pattern transfer. *Langmuir* **17,** 228–233 (2001).
3. Dressick, W. J. & Calvert, J. M. Patterning of self-assembled films using lithographic exposure tools. *Jpn. J. Appl. Phys.* **32,** 5829–5839 (1993).
4. Kuther, J. & Tremel, W. Template induced crystallization of biominerals on self-assembled monolayers of alkylthiols. *Thin Solid Films* **327–329,** 554–558 (1998).
5. Halik, M. *et al.* Low-voltage organic transistors with an amorphous molecular gate dielectric. *Nature* **431,** 963–966 (2004).
6. Krishnamoorthy, A., Chanda, K., Murarka, S. P., Ramanath, G. & Ryan, J. G. Self-assembled near-zero-thickness molecular layers as diffusion barriers for Cu metallization. *Appl. Phys. Lett.* **78,** 2467–2469 (2001).
7. Ganesan, P. G., Gamba, J., Ellis, A., Kane, R. S. & Ramanath, G. Polyelectrolyte nanolayers as diffusion barriers for Cu metallization. *Appl. Phys. Lett.* **83,** 3302–3305 (2003).
8. Ganesan, P. G., Singh, A. P. & Ramanath, G. Diffusion barrier properties of carboxyl- and amine-terminated molecular nanolayers. *Appl. Phys. Lett.* **85,** 579–581 (2004).
9. Ferguson, G. S., Chaudhury, M. K., Sigal, G. B. & Whitesides, G. M. Contact adhesion of thin gold films on elastomeric supports: Cold welding under ambient conditions. *Science* **253,** 776–778 (1991).
10. Ramanath, G. *et al.* Self-assembled subnanolayers as interfacial adhesion enhancers and diffusion barriers for integrated circuits. *Appl. Phys. Lett.* **83,** 383–385 (2003).
11. Maidenberg, D. A., Volksen, W., Miller, R. D. & Dauskardt, R. H. Toughening of nanoporous glasses using porogen residuals. *Nature Mater.* **3,** 464–469 (2004).
12. Neves, B. R. A., Salmon, M. E., Russell, P. E. & Troughton, E. B. J. Thermal stability study of self-assembled monolayers on mica. *Langmuir* **16,** 2409–2412 (2000).
13. Ishida, T. *et al.* Annealing effect of self-assembled monolayers generated from terphenyl derivatized thiols on Au(111). *Langmuir* **18,** 83–92 (2002).
14. Senkevich, J. J., Yang, G. R. & Lu, T. M. Thermal stability of mercaptan terminated self-assembled multilayer films on SiO_2 surface. *Colloids Surf. A* **207,** 139–145 (2002).
15. Gandikota, S. *et al.* Adhesion studies of CVD copper metallization. *Microelectron. Eng.* **50,** 547–553 (2000).
16. Litteken, C. S. & Dauskardt, R. H. Adhesion of polymer thin-films and patterned lines. *Int. J. Fract.* **120,** 475–485 (2003).
17. Snodgrass, J. M., Pantelidis, D., Jenkins, M. L., Bravman, J. C. & Dauskardt, R. H. Subcritical debonding of polymer/silica interfaces under monotonic and cyclic loading. *Acta Mater.* **50,** 2395–2411 (2002).
18. Dauskardt, R. H., Lane, M., Ma, Q. & Krishna, N. Adhesion and debonding of multi-layer thin film structures. *Eng. Fract. Mech.* **61,** 141–162 (1998).
19. Lane, M., Dauskardt, R. H., Krishna, N. & Hashim, I. Adhesion and reliability of copper interconnects with Ta and TaN barrier layers. *J. Mater. Res.* **15,** 203–211 (2000).
20. Evans, A. G., Rühle, M., Dalgleish, B. J. & Charalambides, P. G. The fracture energy of biomaterial interfaces. *Metall. Trans. A* **21,** 2419–2429 (1990).
21. Lane, M. Interface fracture. *Annu. Rev. Mater. Res.* **33,** 29–54 (2003).
22. Hu, M. *et al.* Effect of interfacial interactions on the initial growth of Cu on clean SiO_2 and 3-mercaptopropyltrimethoxysilane-modified SiO_2 substrates. *J. Vac. Sci. Technol. A* **20,** 589–596 (2002).
23. Teo, M., Kim, J., Wong, P. C., Wong, K. C. & Mitchell, K. A. R. Investigations of interfaces formed between bis-1,2-(triethoxysilyl)ethane (BTSE) and aluminum after different Forest Product Laboratory pre-treatment times. *Appl. Surf. Sci.* **221,** 340–348 (2004).
24. D'souza, A. S. & Pantano, C. G. Hydroxylation and dehydroxylation behavior of silica glass fracture surfaces. *J. Am. Ceram. Soc.* **85,** 1499–1504 (2002).
25. Ulman, A. *An Introduction to Ultrathin Organic Films, from Langmuir-Blodgett to Self-Assembly* 1st edn (Academic, San Diego, 1991).
26. Lane, M., Dauskardt, R. H., Vainchtein, A. & Gao, H. Plasticity contributions to interface adhesion in thin-film interconnect structures. *J. Mater. Res.* **15,** 2758–2769 (2000).
27. Guyer, E. P. & Dauskardt, R. H. Fracture of nanoporous thin-film glasses. *Nature Mater.* **3,** 53–57 (2004).
28. Wu, X., Vargas, M. C., Nayak, S., Lotrich, V. & Scoles, G. Towards extending the applicability of density functional theory to weakly bound systems. *J. Chem. Phys.* **115,** 8748–8757 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the US National Science Foundation, the US-Israel Binational Science Foundation, an Alexander von Humboldt fellowship, and New York state through the Interconnect Focus Center.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.R. (ramanath@rpi.edu).

METHODS

MNL formation. MPTMS nanolayers were adsorbed on n-type device quality Si(001) wafers with an 85-nm-thick thermal oxide cap layer, as described elsewhere[10]. The MPTMS MNL thickness was measured by ellipsometry, and verified using variable take-off angle XPS (Supplementary Fig. 2).

Adhesion test sample preparation. After MPTMS MNL formation, the samples were divided into two sets. One set of samples was used as-is for further processing, while the second set was annealed in air for 10 min at 100 °C. XPS showed no observable oxidation of the thiol termini upon this pre-annealing treatment. Thereafter, both sets were subjected to similar processing. Layers of 50-nm-thick Cu and 150-nm-thick Ta were sputter-deposited successively on the MPTMS MNL in a CVC d.c. magnetron tool, without vacuum break. We annealed selected samples at predetermined temperatures of $100\,^{\circ}\mathrm{C} \leq T_{\mathrm{anneal}} \leq 700\,^{\circ}\mathrm{C}$ for 30 min in a 5×10^{-7} torr vacuum furnace to study the effect of annealing on interface toughness. Samples without any annealing treatments were also investigated for comparison. All the samples were glued to a dummy Si wafer using a room-temperature curing epoxy. Following annealing treatments (or the lack thereof), we prepared four-point bend test structures in the configuration illustrated in Fig. 1 inset, using procedures described in detail elsewhere[10]. The Ta layer was sputter-deposited over Cu in order to preclude failure at the Cu/epoxy interface, which is weak. Representative load–displacement curves are shown in Supplementary Fig. 1.

Four-point bend testing and XPS of fracture surfaces. The Si(001)/epoxy/Ta/Cu/MPTMS/SiO_2/Si(001) structures were mounted in the four-point bend test configuration and strained at a rate of 0.01 μm s^{-1} using a piezoelectric actuator in a high-stiffness micromechanical test system. The interface toughness, G, was determined using the method of ref. 18. Each data point in Fig. 1 is an average of the interface toughness determined from at least six separate load–displacement curves measured for samples prepared and tested at different times.

The interface chemistry was studied by examining the fracture surfaces using XPS; instrument details are provided in Supplementary Fig. 2. The fracture surfaces were transferred to the XPS chamber for analysis within 5 min after fracture to minimize oxidation.

Theoretical simulations. Theoretical simulations of molecular stretching and failure were carried out using a density functional theory method based on the B3LYP gradient corrected functional, which is well-suited to accurately capture long range interactions such as hydrogen bonding[28]. We used 6-31+G* basis sets and the Gaussian 03 code for our calculations. The silica surface was described using a cluster model with the dangling bonds saturated by hydrogen atoms. The copper surface was also modelled using a cluster approach, the details of which, and the comparison between results for different cluster sizes, will be published elsewhere. Our calculations consider only one molecule and do not capture intermolecular interactions, or the effect of moisture and other environmental factors. The location of fracture was determined from an analysis of the changes in rate and sign of C–C, Si–O and Cu–S bond lengths with respect to Cu–SiO_2 separation (Supplementary Fig. 3). The system energy at the point of fracture was taken to be the interface toughness.

LETTERS

Survival times of anomalous melt inclusions from element diffusion in olivine and chromite

C. Spandler[1]†, H. St C. O'Neill[1] & V. S. Kamenetsky[2]

The chemical composition of basaltic magma erupted at the Earth's surface is the end product of a complex series of processes, beginning with partial melting and melt extraction from a mantle source and ending with fractional crystallization and crustal assimilation at lower pressures. It has been proposed that studying inclusions of melt trapped in early crystallizing phenocrysts such as Mg-rich olivine and chromite may help petrologists to see beyond the later-stage processes and back to the origin of the partial melts in the mantle[1,2]. Melt inclusion suites often span a much greater compositional range than associated erupted lavas, and a significant minority of inclusions carry distinct compositions that have been claimed to sample melts from earlier stages of melt production, preserving separate contributions from mantle heterogeneities[1–4]. This hypothesis is underpinned by the assumption that melt inclusions, once trapped, remain chemically isolated from the external magma for all elements except those that are compatible in the host minerals[1,2]. Here we show that the fluxes of rare-earth elements through olivine and chromite by lattice diffusion are sufficiently rapid at magmatic temperatures to re-equilibrate completely the rare-earth-element patterns of trapped melt inclusions in times that are short compared to those estimated for the production and ascent of mantle-derived magma[5,6] or for magma residence in the crust[7]. Phenocryst-hosted melt inclusions with anomalous trace-element signatures must therefore form shortly before magma eruption and cooling. We conclude that the assumption of chemical isolation of incompatible elements in olivine- and chromite-hosted melt inclusions[1,2] is not valid, and we call for re-evaluation of the popular interpretation that anomalous melt inclusions represent preserved samples of unmodified mantle melts.

The ability of a host crystal to isolate and preserve a melt inclusion with a different trace-element signature to the external magma (that is, an anomalous melt inclusion) depends on a negligible flux of the trace elements through the host crystal by lattice diffusion. The process has been modelled mathematically by Qin *et al.*[8], who showed that the isolation of the melt inclusion with respect to a trace element M depends on the diffusion coefficient of M and the partition coefficient between melt and crystal, as well as the dimensions of the inclusion and its host crystal. It has been assumed, often implicitly but sometimes explicitly[1,2,8,9], that the partition coefficients and diffusion coefficients of incompatible trace elements like the rare-earth elements (REEs) in the early crystallizing phases olivine and chromite are so low that isolation is effective: that is, anomalous inclusions are preserved over timescales that are long compared to those expected for magma extraction, ascent and residence in the crust. Although it is true that the concentrations of nearly all the incompatible trace elements in olivine and chromite are very low[10], there are no data on the diffusion coefficients, and thus this most important assumption has never been tested. We have conducted multi-component chemical diffusion experiments on melt-inclusion-hosting olivine and chromite crystals in order to evaluate directly the degree of melt inclusion re-equilibration with an external melt. These experiments also enable us simultaneously to determine the diffusion and partition coefficients for REEs and other elements that are needed for calculating re-equilibration timescales from the mathematical model.

For the olivine experiments, pristine olivine crystals (1–2 mm in size) of composition 89–91% forsterite and containing melt inclusions were hand-picked from a sample of high-Mg mid-ocean-ridge basalt dredged from the southern Mid-Atlantic Ridge (sample AG32-4-68 of ref. 11). The inclusions share similar compositions (Fig. 1) that are typical of primitive high-Mg mid-ocean-ridge basalt liquids[12] (Supplementary Information). Trapping temperatures of 1,230–1,280 °C were recorded for around 50 melt inclusions from homogenization on a heating stage[12]. The host olivines have low REE contents that are consistent with values predicted from relevant olivine/melt partition coefficients[13,14]. A glass of basaltic composition calculated to be in equilibrium with the olivine (90% forsterite) at 1,300 °C was synthesized with ~400 p.p.m. of the REEs Pr, Eu, Tb, Ho and Lu (Supplementary Information). The olivine crystals were combined with the powdered synthetic glass and placed in Re buckets in a gas-mixing furnace and run at 1,300 °C, sufficient to ensure complete remelting of the inclusions, with oxygen fugacity set to $10^{-9.7}$ bar using a flowing CO/CO_2 gas mixture. Run durations were 1, 5 and 25 days, after which the samples were quenched rapidly by dropping into water. Olivine grains were polished to expose the inclusions for analysis of major elements (by energy dispersive spectrometry) and of REEs and selected other trace elements (by laser-ablation inductively coupled plasma mass spectrometry, ICP-MS), with a number of elemental profiles across olivine grains obtained to determine both diffusion coefficients and crystal/melt partition coefficients. The diffusion coefficients are averages from randomly oriented crystallographic sections, suitable for applying to the model of Qin *et al.*[8], which assumes isotropic diffusion. Similar experimental techniques were used for inclusion-hosting chromite grains from the Stillwater Complex[15], except the experimental duration was 7 days at 1,450 °C, to ensure remelting and homogenization of the melt inclusions[15], and to produce easily measurable diffusion profiles. The experimental and analytical techniques are more fully described in the Supplementary Information.

The REE contents in the melt inclusions from the one-day olivine experiment are indistinguishable from the original compositions (Fig. 1a), indicating that diffusive re-equilibration of these inclusions from the synthetic external melt was insignificant. However, sections of the olivine up to 50 μm from the olivine/external melt interface have elevated Lu, Ho and Tb contents. Larger inclusions from the

[1]Research School of Earth Sciences, Australian National University, Canberra 0200, Australia. [2]ARC Centre of Excellence in Ore Deposits and School of Earth Sciences, University of Tasmania, Hobart 7001, Australia. †Present address: Institute of Geological Sciences, University of Bern, Bern 3012, Switzerland.

five-day run were also mostly unaffected, but many smaller (<40 μm) inclusions were enriched in Lu, Ho, Tb and Eu (Fig. 1b), and olivine grains had elevated Lu, Ho, Tb and Eu contents to 100 μm from the external melt interface. After 25 days, 21 out of a total of 22 analysed inclusions have modified REE patterns (Fig. 1c). Inclusions with the most extreme REE patterns tend to be small, located within 50 μm of an olivine grain boundary, and are measurably enriched in Pr as well as Lu, Ho, Tb and Eu. They are also depleted in the heavy REEs not added to the external melt (Er, Tm, Yb). Elevated levels of the doped REEs were also found in the olivine crystals, decreasing from the melt/olivine interface towards the interior of the crystal along diffusion profiles (Fig. 2; Supplementary Fig. 1). Similar results were obtained from the chromite experiment. The REE content of chromite has usually been considered insignificant[10], but diffusion of REEs at 1,450 °C is sufficiently rapid that modification of chromite-hosted melt inclusions is observed after only 7 days (Fig. 1d).

All melt inclusions with altered REE patterns show a monotonic increase in the degree of enrichment of the doped REE with atomic number, as do their host olivine or chromite crystals. Lutetium is the most extensively enriched element, followed by Ho, Tb and Eu, with elevated Pr contents only found in inclusions with the most modified REE patterns. Because the external melt was doped with similar contents of the selected REEs (Supplementary Table 1), this fractionation of the REEs in the melt inclusions and olivine cannot be due to physical mixing between the melt inclusions and the external melt, for example along micro-cracks. Furthermore, optical and chemical examination of the olivine and chromite grains indicates that the grains did not undergo dissolution or crystallization during the experiments. The enrichment and fractionation of REEs is attributed solely to diffusion of these elements from the external melt through the olivine and chromite and thence into their melt inclusions. The depletion of Er, Tm and Yb in some of the inclusions (Fig. 1c) is due to diffusion of these elements out of the inclusions, as the external melt is initially devoid of these elements. The general inverse relationship between melt inclusion size and degree of element enrichment (Fig. 1b, c) is expected, because smaller inclusions are more susceptible to diffusive re-equilibration owing to their greater surface area to volume ratio[8,9].

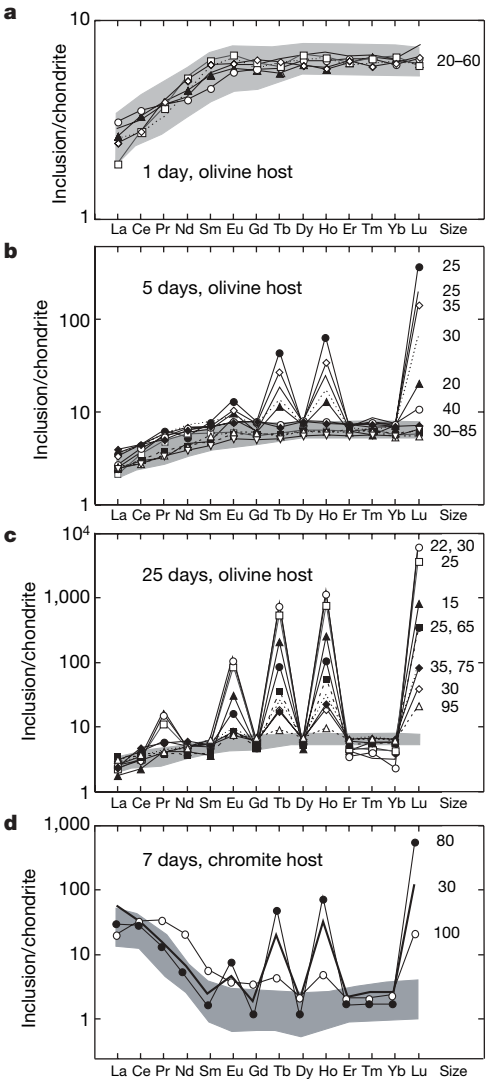


Figure 1 | Chondrite-normalized REE plots of olivine- and chromite-hosted melt inclusions. The grey fields represent the compositional range of normal melt inclusions after homogenization. **a**–**c**, Representative olivine-hosted melt inclusions from the 1-day (**a**), 5-day (**b**) and 25-day (**c**) experiments. **d**, Chromite-hosted melt inclusions from a 7-day experiment. Inclusion size is in μm.

Diffusion coefficients and partition coefficients were obtained by least-squares fitting of the analytical traverses (Fig. 2) (see Supplementary Information). Mean values are given in Table 1 together with observed standard deviations. Our measured diffusion coefficients for Ca and Ni in olivine are in excellent agreement with previous work using quite different experimental approaches[16,17] (see Table 1). A novel feature of our experiments is that the chemical potentials of all relevant major-element oxide components are controlled by the presence of the silicate melt, and moreover are controlled at values appropriate to natural magmas. Previous investigations of diffusion in olivine and other silicates have until recently only controlled oxygen fugacity among chemical potentials, despite the expectation that point-defect concentrations, which determine diffusion rates, depend on other chemical potentials such as silica activity[18,19]. New results on the effect of H_2O on interdiffusion of Mg-Fe^{2+} in olivine demonstrate this clearly[20].

Our reported values of the partition coefficients (K_M) for the REEs between olivine and melt (Table 1) are similar to previously reported values[13,14] (see Supplementary Information). Values of the diffusion coefficients (D_M) of the REEs are identical within experimental error for all REEs in both olivine and chromite, as previously found for garnet, feldspar and apatite, but in contrast to diopside and zircon[21]; the reason for this dichotomy is not known. However, values of the REE diffusion coefficients are two to four orders of magnitude faster

Table 1 | Calculated parameters for olivine and chromite

Element	log D_M ($m^2 s^{-1}$)	$\pm\sigma$	K_M	$\pm\sigma$
Olivine (Mg# ≈ 90; T = 1,300 °C; $f_{O_2} = 10^{-9.7}$ bar)				
Ca	−15.72	0.23	0.015	3×10^{-3}
Ca*	−15.53			
Ca†	−15.39 to −15.25‡			
Ti	−14.99	0.24	0.09	0.03
Ni	−15.20	0.23	20	2
Ni*	−15.21 to −14.82‡			
Pr	−15.17	0.27	17×10^{-5}	3×10^{-5}
Eu	−14.89	0.18	4×10^{-4}	1×10^{-4}
Tb	−14.83	0.14	26×10^{-4}	5×10^{-4}
Ho	−14.86	0.17	7×10^{-3}	1×10^{-3}
Lu	−14.95	0.17	0.033	0.007
Chromite (T = 1,450 °C; $f_{O_2} = 10^{-6.8}$ bar)				
Ni	−14.52	0.57	38.3	15.3
Pr	−14.49	0.47	8×10^{-5}	3×10^{-5}
Eu	−14.61	0.07	20×10^{-5}	5×10^{-5}
Tb	−14.43	0.24	9×10^{-4}	3×10^{-4}
Ho	−14.46	0.26	21×10^{-4}	8×10^{-4}
Lu	−14.50	0.29	76×10^{-4}	26×10^{-4}

Shown are calculated diffusion coefficients (D_M) and partition coefficients (K_M); for each value, uncertainties ($\pm\sigma$) are given in the following column. Note the similarity between Ca and Ni diffusion coefficients calculated from this study and those reported by ref. 16 and ref. 17. f_{O_2}, oxygen fugacity.
* Data from ref. 17 at 1,300 °C and f_{O_2} of $\sim 10^{-9}$ bar.
† Data from ref. 16 at 1,300 °C and f_{O_2} of $\sim 10^{-10}$ bar.
‡ Range is that covered by different crystallographic orientations or varying olivine composition.

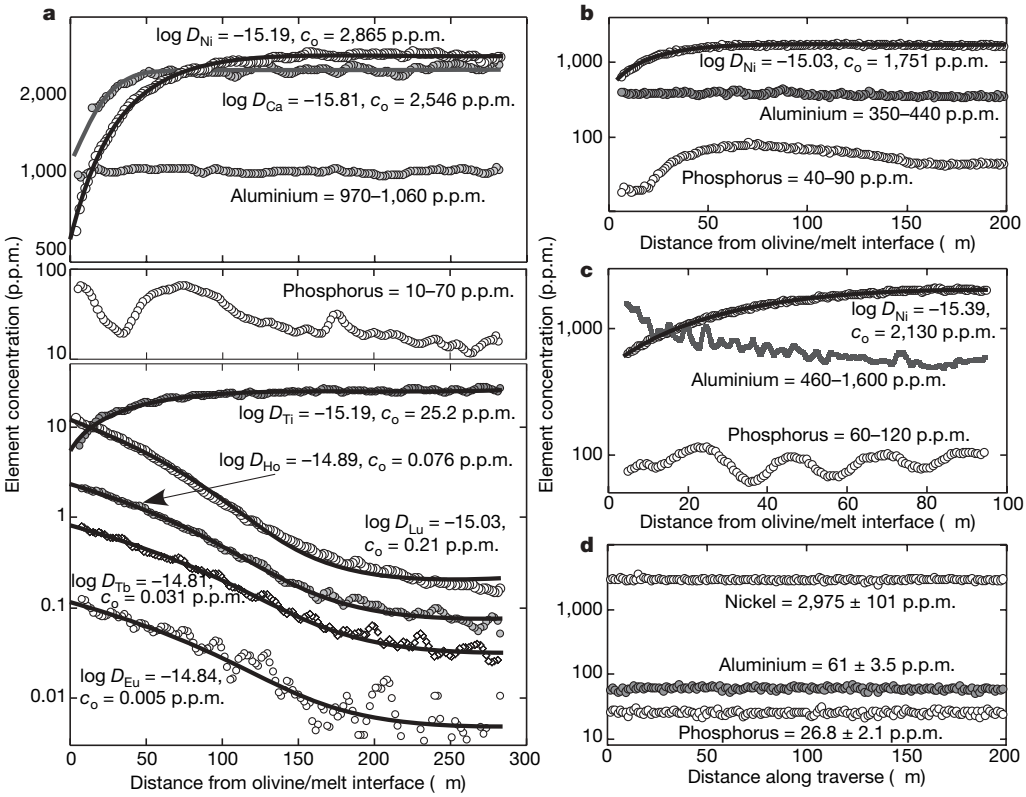


Figure 2 | Measured element diffusion profiles from analytical traverses (obtained by laser ablation ICP-MS) across olivine grains. Solid lines represent the fitted diffusion profiles. **a**, Profiles for Ni, Ca, Al, P, Ti, Lu, Ho, Tb and Eu across olivine from the 25-day experiment. The increased scatter in Eu and Tb values at 150–300 μm from the olivine/melt interface reflect low element concentrations that approach the analytical detection limits. **b**, Profiles for Ni, Al and P across olivine from the 5-day experiment. **c**, Profiles for Ni, Al and P across a second olivine grain from the 25-day experiment. **d**, Profile for Ni, Al and P across the interior of a homogenous olivine from a spinel-lherzolite xenolith (Anakies 9A), for comparison. As P was not present in the melt, any measurable diffusion of P would result in very low levels of this element in the olivine at equilibrium (≤1 p.p.m.). Note the total lack of diffusion of P and the preservation of igneous zoning in the Al and P profiles. Olivines show a variety of original Ni concentrations, but values of D_{Ni} are identical within expected uncertainty. Note the homogeneity of the Anakies 9a olivine compared with the phenocrystal olivines used in the experiments. Details of the data fitting and analysis are presented in the Supplementary Information. The logarithms of diffusion coefficient values (D_M in $m^2 s^{-1}$) and either the original concentration (c_o) or, for non-diffusing elements, concentration ranges, are indicated.

in olivine than in these aforementioned minerals at the same temperature[21], and are remarkably similar to the diffusion coefficients for Mg and Fe^{2+} (ref. 17), the major divalent cations in olivine. The only study addressing the substitution mechanisms of large trivalent cations into olivine found that the substitution of Sc is coupled to vacancies (vac) to maintain charge balance, producing the end-member $(Sc_{2/3}vac_{1/3})_2SiO_4$ (ref. 22). This study found no evidence for Sc coupling with Al (that is, as an $MScAlO_4$ component, where M = Mg or Fe^{2+}). If REEs substitute similarly, then charge balance requires that the diffusion of REEs through the olivine lattice occurs by the mechanism:

$$4/3\,REE^{3+} + M_2SiO_4 = 2\,M^{2+} + (REE_{2/3}vac_{1/3})_2SiO_4$$

The similarity of the observed diffusion coefficients of the REEs to those of Mg and Fe suggest that the latter are the rate-controlling factors for REE diffusion in olivine, which also accounts for the observation that the values of diffusion coefficients for the REEs are independent of atomic number. In contrast to the fast diffusion of REEs, very slow diffusion of Al and P (two elements that may substitute for slow-diffusing Si in the tetrahedral site of olivine) is evident from the absence of any observable diffusion profiles, and from the variety of original magmatic concentrations and zoning patterns that are preserved (Fig. 2). In some crystals (for example, Fig. 2c), Al is heterogeneous on a scale similar to the resolution of the analytical method (~3 μm), suggesting that diffusion rates of Al in olivine are at least three orders of magnitude slower than those of REEs or divalent cations.

It has been assumed that diffusion does not significantly alter the incompatible trace element content of olivine- or chromite-hosted melt inclusions[1,2,8,9] because partition coefficients for these elements in olivine and chromite are low (this work, and refs 10 and 13). However, REE diffusion in these minerals is sufficiently rapid to counter the effects of low partition coefficients, permitting a significant flux of REEs through the olivine and chromite in times that are short compared to those typically proposed for processes of partial melting, magma transport or crustal residence (10 to 10^5 years)[5–7].

We have used our diffusion coefficients and olivine/melt partition coefficients (Table 1) with the mathematical model of Qin *et al.*[8] to calculate timescales of REE re-equilibration between external melt and a 50-μm spherical melt inclusion trapped within a 1-mm olivine grain. Our results show that a melt inclusion will undergo 90% re-equilibration in less than 3 years for the heavy REEs, in around 10 years for the middle REEs and in 20–100 years for the light REEs (Fig. 3). These limiting times may be shortened further in hydrous magmas if the effect of H_2O, not present in our experiments, is included[20]. Recent estimates for the transit times of basalt from its mantle source through the crust range from 30 to $>10^4$ years, from measurements of U-series disequilibria[5] or geophysical modelling of melt extraction processes[6]. Residence times of basaltic magma and associated crystals in the lower crust are also expected to be decades to thousands of years[7]. Our data affirm that even if the shorter timescales for these processes are accepted, the REE patterns of anomalous melt inclusions trapped in olivine phenocrysts in the mantle or lower crust will be extensively modified by lattice diffusion before transport to the Earth's surface. These constraints require olivine-hosted melt

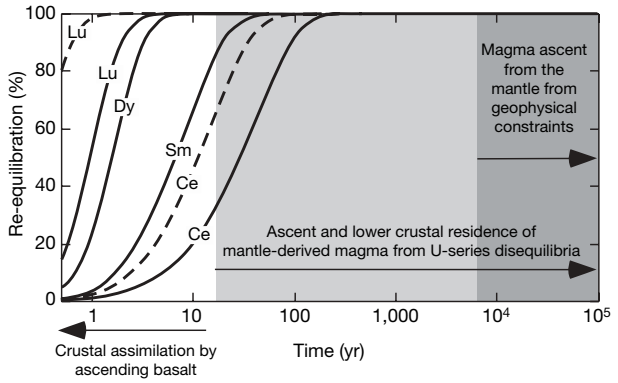


Figure 3 | Modelled re-equilibration times for REEs between a melt inclusion in an olivine grain and an external melt at 1,300 °C. Solid curves, a 50-μm melt inclusion in a 1-mm grain; dashed curve for Lu, a 50-μm melt inclusion in a 0.5-mm grain; dashed curve for Ce, a 30-μm melt inclusion in a 1-mm grain. Re-equilibration was calculated from the equations of ref. 8 using a diffusion coefficient of $10^{-14.9}$ m^2 s^{-1} for all REEs and relevant REE olivine/melt partition coefficients (see Supplementary Fig. 2). Shown for comparison are timescales for basalt transfer from the mantle from U-series disequilibria[5] and from geophysical constraints[6], and residence of basalt and associated crystals in the lower crust[7]—all of which are longer than the time needed for re-equilibration of REE in the melt inclusion. By contrast, timescales of assimilation of crustal material by ascending basalt are relatively short[27].

inclusions with anomalous REE concentrations to be trapped at upper crustal levels shortly (<20 years) before eruption. Melt inclusions trapped in mantle xenoliths from kimberlites, which are transported to the surface in hours to days[23], would be exceptions. This result is consistent with studies that demonstrate some melt inclusion suites to form at relatively low pressures[2,3] or immediately preceding eruption and cooling[24], and we propose that melt inclusion entrapment in phenocrysts is likely to be a late-stage process in general.

Studies of exposed crustal sequences[25,26] and xenocrysts in volcanic rocks[27] indicate that primary mantle melts often experience some degree of reaction and modification during transit through the crust. The fast rates of REE diffusion in olivine and chromite reported here confirm that the isolation of inclusions by these phases may not be sufficient to preserve them against such modification, calling into question the view that anomalous melt inclusions faithfully record signatures of primary melts from the mantle[1–4]. Our results are compatible with alternative proposals that anomalous melt inclusions form either during localized assimilation of crustal rocks—for example, at the margins of a magma chamber or during ascent in a magma conduit[28,29]—or during shallow-level magma mixing[30]. This scenario is consistent with constraints from divalent cation diffusion profiles in olivine xenocrysts, which demonstrate that crustal assimilation by ascending basaltic magmas occurs shortly before eruption (<10 years) (ref. 27). Melt inclusion studies may therefore provide important insights into crustal modification of mantle-derived magmas[28].

Received 20 June 2006; accepted 16 March 2007.

1. Schiano, P. Primitive mantle magmas recorded as silicate melt inclusions in igneous minerals. *Earth-Sci. Rev.* **63,** 121–144 (2003).
2. Sobolev, A. V. Melt inclusions in minerals as a source of principal petrological information. *Petrology* **4,** 209–220 (1996).
3. Sobolev, A. V., Hofmann, A. W. & Nikogosian, I. K. Recycled oceanic crust observed in 'ghost plagioclase' within the source of Mauna Loa lava. *Nature* **404,** 986–990 (2000).
4. Ren, Z.-Y., Ingle, S., Takahashi, E., Hirano, N. & Hirata, T. The chemical structure of the Hawaiian mantle plume. *Nature* **436,** 837–840 (2005).
5. Condomines, M., Gauthier, P.-J. & Sigmarsson, O. Timescales of magma chamber processes and dating of young volcanic rocks. *Rev. Miner. Geochem.* **52,** 125–174 (2003).
6. Faul, U. H. Melt retention and segregation beneath mid-ocean ridges. *Nature* **410,** 920–923 (2001).
7. Reid, M. R. in *Treatise on Geochemistry* Vol. 3, *The Crust* (ed. Rudnick, R. L.) 167–193 (Elsevier Science, Amsterdam, 2003).
8. Qin, Z., Lu, F., Anderson, A. T. Jr., Diffusive reequilibration of melt and fluid inclusions. *Am. Mineral.* **77,** 565–576 (1992).
9. Cottrell, E., Spiegelman, M. & Langmuir, C. H. Consequences of diffusive reequilibration for the interpretation of melt inclusions. *Geochem. Geophys. Geosyst.* **3,** 1026 (2002).
10. Witt-Eickschen, G. & O'Neill, H. St C. The effect of temperature on the equilibrium distribution of trace elements between clinopyroxene, orthopyroxene, olivine and spinel in upper mantle peridotite. *Chem. Geol.* **221,** 65–101 (2005).
11. Le Roex, A. P., Dick, H. J. B., Gulen, L., Reid, A. M. & Erlank, A. J. Local and regional heterogeneity in MORB from the Mid-Atlantic Ridge between 54.5°S and 51°S: Evidence for geochemical enrichment. *Geochim. Cosmochim. Acta* **51,** 541–555 (1987).
12. Kamenetsky, V. Methodology for the study of melt inclusions in Cr-spinel, and implications for parental melts of MORB from FAMOUS area. *Earth Planet. Sci. Lett.* **142,** 479–486 (1996).
13. Kennedy, A. K., Lofgren, G. E. & Wasserburg, G. J. An experimental study of trace element partitioning between olivine, orthopyroxene and melt in chondrules: Equilibrium values and kinetic effects. *Earth Planet. Sci. Lett.* **115,** 177–195 (1993).
14. McKay, G. A. Crystal/liquid partitioning of REE in basaltic systems: Extreme fractionation of REE in olivine. *Geochim. Cosmochim. Acta* **50,** 69–79 (1986).
15. Spandler, C. J., Mavrogenes, J. A. & Arculus, R. J. The origin of chromitites in layered intrusions: Evidence from chromite-hosted melt inclusions from the Stillwater Complex. *Geology* **11,** 893–896 (2005).
16. Jurewicz, A. J. G. & Watson, E. B. Cations in olivine, Part 2: Diffusion in olivine xenocrysts, with applications to petrology and mineral physics. *Contrib. Mineral. Petrol.* **99,** 186–201 (1988).
17. Petry, C., Chakraborty, S. & Palme, H. Experimental determination of Ni diffusion coefficients in olivine and their dependence on temperature, composition, oxygen fugacity, and crystallographic orientation. *Geochim. Cosmochim. Acta* **68,** 4179–4188 (2004).
18. Stocker, R. L. & Smyth, D. M. Effect of enstatite activity and oxygen partial pressure on the point-defect chemistry of olivine. *Phys. Earth Planet. Inter.* **16,** 145–156 (1978).
19. Nakamura, A. & Schmalzried, H. On the nonstoichiometry and point defects of olivine. *Phys. Chem. Miner.* **10,** 27–37 (1983).
20. Hier-Majumder, S., Anderson, I. M. & Kohlstedt, D. L. Influence of protons on Fe-Mg interdiffusion in olivine. *J. Geophys. Res.* **110,** doi:2004JB003292 (2005).
21. Cherniak, D. J. REE diffusion in feldspar. *Chem. Geol.* **193,** 25–41 (2003).
22. Nielsen, R. L., Gallahan, W. E. & Newberger, F. Experimentally determined mineral-melt partition coefficients for Sc, Y, REE for olivine, orthopyroxene, pigeonite, magnetite and ilmenite. *Contrib. Mineral. Petrol.* **110,** 488–499 (1992).
23. Kelley, S. P. & Wartho, J. A. Rapid kimberlite ascent and the significance of Ar-Ar ages in xenolith phlogopites. *Science* **289,** 609–611 (2000).
24. Danyushevsky, L. V., Sokolov, S. & Falloon, T. J. Melt inclusions in olivine phenocrysts: Using diffusive re-equilibration to determine the cooling history of a crystal, with implications for the origin of olivine-phyric volcanic rocks. *J. Petrol.* **43,** 1651–1671 (2002).
25. Bedard, J. H. Oceanic crust as a reactive filter — synkinematic intrusion, hybridization and assimilation in an ophiolitic magma chamber, western Newfoundland. *Geology* **21,** 77–80 (1993).
26. Coogan, L. A. Contaminating the lower crust in the Oman ophiolite. *Geology* **31,** 1065–1068 (2003).
27. Costa, F. & Dungan, M. Short time scales of magmatic assimilation from diffusion modelling of multiple elements in olivine. *Geology* **33,** 837–840 (2005).
28. Danyushevsky, L. V., Leslie, R. A. J., Crawford, A. J. & Durance, P. Melt inclusions in primitive olivine phenocrysts: The role of localized reaction processes in the origin of anomalous compositions. *J. Petrol.* **45,** 2531–2553 (2004).
29. Yaxley, G. M., Kamenetsky, V. S., Kamenetsky, M., Norman, M. D. & Francis, D. Origins of compositional heterogeneity in olivine-hosted melt inclusions from the Baffin Island picrites. *Contrib. Mineral. Petrol.* **148,** 426–442 (2004).
30. Perugini, D., Petrelli, M. & Poli, G. Diffusive fractionation of trace elements by chaotic mixing of magmas. *Earth Planet. Sci. Lett.* **243,** 669–680 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements M. Shelley, A. Norris and D. Scott are thanked for their help with the laser ablation ICP-MS analyses, electron microprobe analyses, and experimental set-up, respectively. This work was supported by an Australian Research Council Discovery Grant (to H.StC.O'N.).

Author Contributions C.S. prepared and performed the experiments, and H.StC.O'N. fitted the analytical data to obtain diffusion coefficients. C.S. and H.StC.O'N. conducted the microprobe and laser-ablation ICP MS analyses. V.K. characterized and supplied the sample of melt inclusion-bearing olivine phenocrysts. C.S. and H.StC.O'N. co-wrote the paper. All authors discussed the results and commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.S. (spandler@geo.unibe.ch) or H.StC.O'N. (hugh.oneill@anu.edu.au).

LETTERS

First insights into the biodiversity and biogeography of the Southern Ocean deep sea

Angelika Brandt[1], Andrew J. Gooday[2], Simone N. Brandão[1], Saskia Brix[1], Wiebke Brökeland[1], Tomas Cedhagen[3], Madhumita Choudhury[1], Nils Cornelius[2], Bruno Danis[4], Ilse De Mesel[5], Robert J. Diaz[6], David C. Gillan[7], Brigitte Ebbe[8], John A. Howe[9], Dorte Janussen[10], Stefanie Kaiser[1], Katrin Linse[11], Marina Malyutina[12], Jan Pawlowski[13], Michael Raupach[14] & Ann Vanreusel[5]

Shallow marine benthic communities around Antarctica show high levels of endemism, gigantism, slow growth, longevity and late maturity, as well as adaptive radiations that have generated considerable biodiversity in some taxa[1]. The deeper parts of the Southern Ocean exhibit some unique environmental features, including a very deep continental shelf[2] and a weakly stratified water column, and are the source for much of the deep water in the world ocean. These features suggest that deep-sea faunas around the Antarctic may be related both to adjacent shelf communities and to those in other oceans. Unlike shallow-water Antarctic benthic communities, however, little is known about life in this vast deep-sea region[2,3]. Here, we report new data from recent sampling expeditions in the deep Weddell Sea and adjacent areas (748–6,348 m water depth) that reveal high levels of new biodiversity; for example, 674 isopods species, of which 585 were new to science. Bathymetric and biogeographic trends varied between taxa. In groups such as the isopods and polychaetes, slope assemblages included species that have invaded from the shelf. In other taxa, the shelf and slope assemblages were more distinct. Abyssal faunas tended to have stronger links to other oceans, particularly the Atlantic, but mainly in taxa with good dispersal capabilities, such as the Foraminifera. The isopods, ostracods and nematodes, which are poor dispersers, include many species currently known only from the Southern Ocean. Our findings challenge suggestions that deep-sea diversity is depressed in the Southern Ocean and provide a basis for exploring the evolutionary significance of the varied biogeographic patterns observed in this remote environment.

Although animal communities inhabiting shallow marine benthic environments around Antarctica, notably parts of McMurdo Sound and the Peninsula region, are well known, there have been few studies of the deep-water faunas in the adjacent Southern Ocean. The ANDEEP (Antarctic benthic deep-sea biodiversity: colonization history and recent community patterns) project was designed to fill this knowledge vacuum[4]. Between 2002 and 2005, we undertook three expeditions in the deep Weddell Sea and adjacent areas aboard the German research vessel *Polarstern*. Biological collections and data on environmental and seafloor characteristics were obtained from diverse settings, including continental slope, rise, abyssal plain, trench floor, channel levees and adjacent to fracture zones, and from water depths between 774 and 6,348 m (Fig. 1; Supplementary Discussion; Supplementary Tables 1, 2; Supplementary Figs 1, 2). This material has greatly improved our knowledge of the biodiversity of benthic communities in the deep Southern Ocean and enabled us to test ideas about large-scale biogeographic and other macroecological patterns[5] among deep-water faunas.

The richness of species inhabiting deep-sea sediments was first demonstrated during the 1960s[6,7]. Since then, scientists have sought to understand the mechanisms that maintain these high levels of benthic diversity at local[8] and regional[9] scales. We analysed species assemblages in ANDEEP samples across a range of taxonomic groups, representing the meiofauna, macrofauna and megafauna (Fig. 2) and found substantial levels of unrecorded biodiversity (see Supplementary Material). The Foraminifera were represented by 158 live species, including 72 monothalamous species, most of them undescribed. The nematodes belonged to typical cosmopolitan deep-sea genera (Supplementary Fig. 3), but more than half of the 57

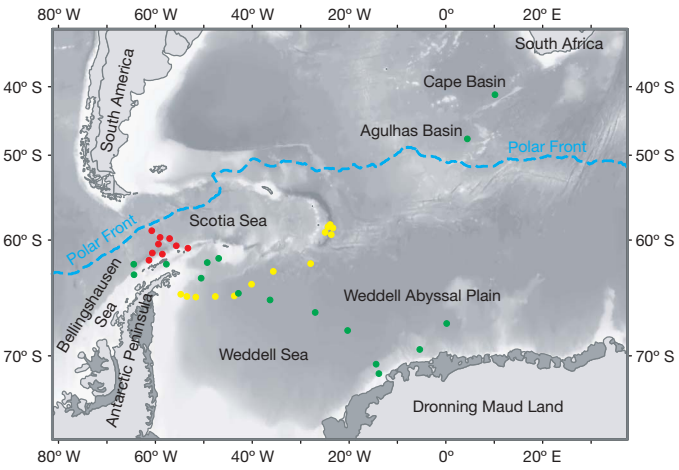


Figure 1 | Station map. Stations from ANDEEP I (2002) red circles, ANDEEP II (2002) yellow circles and ANDEEP III (2005) green circles. The blue line indicates the Polar Front.

[1]Zoological Museum Hamburg, Martin-Luther-King-Platz 3, 20146 Hamburg, Germany. [2]National Oceanography Centre, Southampton, European Way, Southampton SO14 3ZH, UK. [3]Department of Marine Ecology, University of Åarhus, Finlandsgade 14, 8200 Århus N, Denmark. [4]Royal Belgian Institute of Natural Sciences, 29 Rue Vautier, 1000 Brussels, Belgium. [5]Biology Department, Marine Biology Section, Ghent University, Krijgslaan 281/S8, 9000 Ghent, Belgium. [6]Virginia Institute of Marine Science, College of William and Mary, Gloucester Point, Virginia 23062, USA. [7]Marine Biology Laboratory, CP160/15, Université Libre de Bruxelles, 50 Avenue Roosevelt, 1050 Brussels, Belgium. [8]Forschungsinstitut Senckenberg, DZMB-CeDAMar, c/o Forschungsmuseum König, Adenauerallee 160, 53113 Bonn, Germany. [9]Scottish Association for Marine Science, Dunstaffnage Marine Laboratory, Dunbeg, Oban, Argyll PA37 1QA, UK. [10]Forschungs und Naturmuseum Senckenberg, Senckenberganlage 25, 60325 Frankfurt am Main, Germany. [11]British Antarctic Survey, Natural Environmental Research Council, High Cross, Madingley Road, Cambridge CB3 0ET, UK. [12]Institute of Marine Biology, FEB RAS, Palchevskogo, 17, 690041 Vladivostok, Russia. [13]Department of Zoology & Animal Biology, University of Geneva, 30 Quai Ernest Ansermet, CH 1211 Genève 4, Switzerland. [14]Ruhr-Universität Bochum, Universitätstrasse 150, 44780 Bochum, Germany.

species recognized in selected genera were new to science. At least 100 ostracod species were distinguished, and >70% of them were new. Among the macrofauna, the isopods were astonishingly diverse with 674 species identified among >13,000 specimens examined, compared with 371 species reported from the entire Antarctic continental shelf[10]. Most (86%) of the isopod species were undescribed and are presently known only from the Southern Ocean. More than 200 polychaete species were recognized, 81 of them previously unknown. Our samples yielded 160 species of shelled gastropods and bivalves compared with 279 species on the shelf (<1,000 m)[11]. Around 40% of species were confined to abyssal depths. Among the megafauna, 76 species of sponges were recognized, 17 of them were undescribed and 37 new for the Southern Ocean (Supplementary Table 3).

Whether shelf faunas have colonized the deep ocean, or vice versa, is a central issue for understanding faunal patterns in the Southern Ocean. Prior to ANDEEP, most of the information on bathymetric distributions around Antarctica concerned shelf and upper-slope faunas[10]. A comparison of species distributions on continental shelves and slopes around Europe and Antarctica revealed significantly wider depth ranges off Antarctica for bivalves, gastropods, amphipods and decapods but not for polychaetes, asteroids and ophiuroids[12]. The generally higher degree of eurybathy among Antarctic invertebrates was attributed to glacial–interglacial cycles of shelf ice advance and retreat, which periodically eliminated shelf faunas, pushing species into deeper water or causing their extinction. These cycles probably also led to regular pulses of migration in and out of Antarctica[1]. Experimental studies of the pressure and temperature tolerances of the pelagic larvae of the shallow-water Antarctic echinoid *Sterechinus neumayeri* suggested that embryonic stages potentially could penetrate into deep water[13]. The unusually deep shelf around Antarctica could also facilitate faunal exchanges between shelf and deep-sea habitats[2].

ANDEEP data strongly suggest that a combination of emergence and submergence on Antarctic margins has led to an intermingling of species originating in shallow- and deep-water habitats (Fig. 3). The

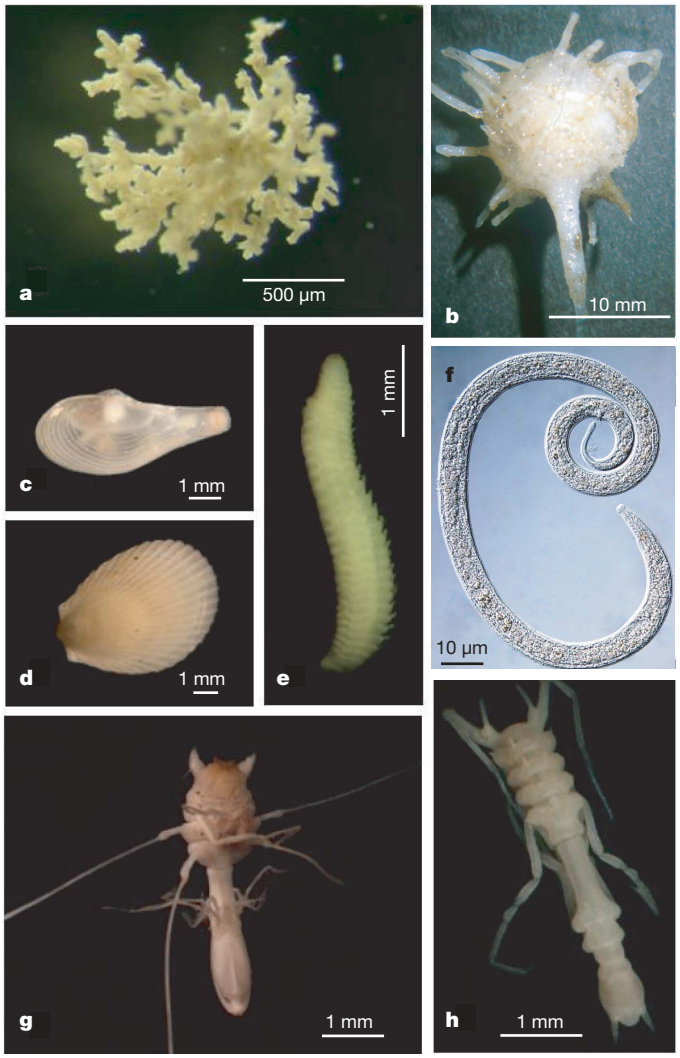


Figure 2 | Selected important ANDEEP taxa. **a**, A new species of komokiacean (*Ipoa* sp. nov.) from the Weddell Sea abyssal plain. **b**, The carnivorous sponge *Chondrocladia* from the Weddell Sea abyssal plain. **c**, Bivalve mollusc *Cuspidaria* sp. from the abyssal Weddell Sea and South African basins. **d**, Bivalve mollusc *Limatula* sp. from the Weddell Sea. **e**, Polychaete *Ophryotrocha* from the Weddell Sea. **f**, Nematode *Molgolaimus* sp. from the Weddell Sea. **h**, Isopod crustacean *Ischnomesus* sp. from the Weddell Sea. **g**, Isopod *Munnopsis* sp. from the western Weddell Sea.

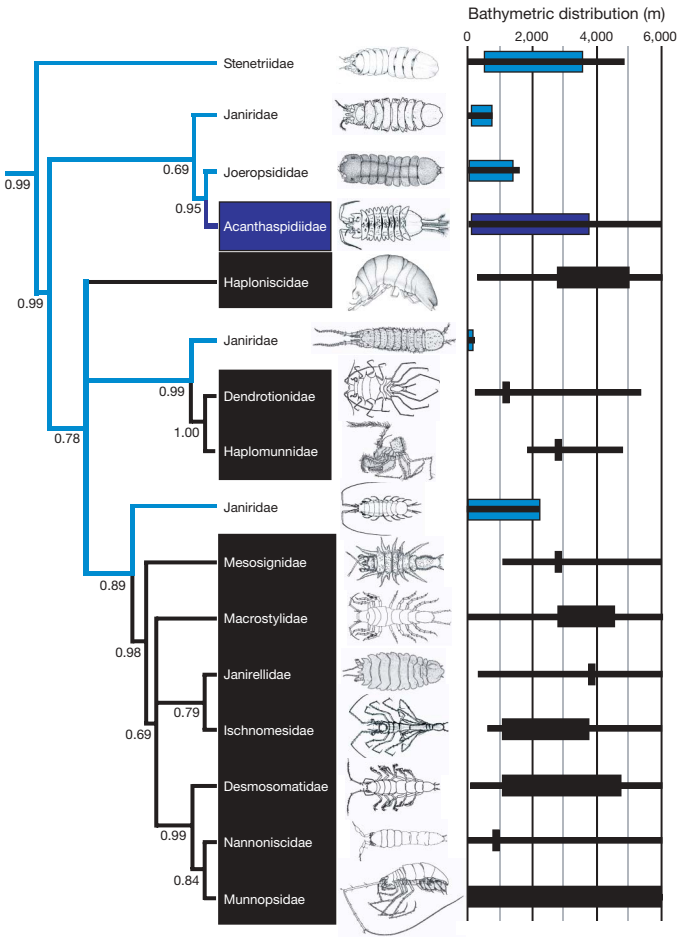


Figure 3 | Isopod phylogenetic relationships and multiple colonizations of the deep sea. Bayesian consensus tree based on the complete 18S rRNA gene (excluding hypervariable regions; numbers on nodes represent statistical support) of 53 species of marine Asellota, and bathymetric distribution of the analysed species (broad lines). Family distribution (thin black lines) is based on ANDEEP data and the world list of isopods (www.nmnh.si.edu./iz/isopod/). Eye-bearing, shallow-water taxa are marked light blue; blind, deep-sea taxa are marked black. The family Acanthaspidiidae includes both shallow-water and deep-sea species (dark blue colour). The scattered distribution of the shallow-water forms within the tree, caused by the polyphyletic nature of the 'Janiridae', indicates a multiple colonization of the deep sea. In at least four cases, asellote families (including the Acanthaspidiidae) have invaded the deep sea. The Munnopidae are most speciose and include seven dominant deep-sea subfamilies. The bathymetric distribution of the analysed species provides evidence of polar submergence within the Stenetriidae, Joeropsididae and some janirid species, whereas polar emergence can be observed within the Acanthaspidiidae (genus *Acanthaspidia*), Dendrotionidae, Ischnomesidae, Desmosomatidae, Nannoniscidae and Munnopsidae.

foraminiferan *Epistominella vitrea* provides a good example of submergence. Populations from 18 to 28 m in McMurdo Sound and from 1,080 m on the opposite side of the continent in the Weddell Sea are genetically identical[14]. This species, which typically resides on the shelf, has a much greater depth range in the Southern Ocean than it does on other continental margins, suggesting a migration from the shelf into deep water. *Epistominella exigua,* the deep-water counterpart of *E. vitrea*, which is reported from depths of 5,000 m in the central Weddell Sea[15] and 314 m on the shelf[16], may have migrated in the opposite direction. Among nematodes, typical deep-sea members of the genus *Acantholaimus* have emerged onto the Southern Ocean shelf[17]. Deep-sea emergence of isopods was postulated by several authors[18], but others have regarded submergence of shelf species into the deep sea as more likely[19]. Probably, both exchange pathways have been followed by different isopod families over evolutionary timescales; for example, submergence by the Acanthaspidiidae, Antarcturidae, Munnidae, Serolidae, Stenetriidae, Paramunnidae and emergence by the Desmosomatidae, Ischnomesidae, Munnopsidae. ANDEEP samples reveal that isopod species typical of the shelf penetrate to 1,500–2,000 m depth. Some of the deep-water antarcturid species have eyes, suggesting that they evolved from shelf species. The abyssal ANDEEP isopods, however, are completely different from those on the shelf, consisting almost entirely (97%) of asellotes. Molecular genetic analyses demonstrate that the remarkable radiation of deep Southern Ocean asellotes encompasses at least four different lineages that have colonized the deep-sea independently (Fig. 3). This group also includes some species that emerge to depths of a few 100s of metres on the shelf[20]. In other crustacean taxa, the occurrence of the thaerocytherid ostracod *Australicythere* cf. *polylyca* at 1,030 m depth in the eastern Weddell Sea may indicate submergence from the Antarctic shelf, whereas three macrocypridid species that occur from shelf to abyssal depths provide an example of emergence in this deep-sea family. Among polychaetes, a eurybathic shelf assemblage extends beyond 2,000 m, at which depth it is replaced by a lower-slope/abyssal-plain assemblage. Although the depth ranges themselves are not unusual for polychaetes, the downslope species turnover is different from temperate regions. Among molluscs, there is evidence for the emergence of species evolved from the deep-sea bivalve *Limopsis tenella* onto shelf. However, abyssal Southern Ocean buccinoid gastropods share no genera with sublittoral and bathyal faunas.

Although the Polar Front (Fig. 1) limits the distribution of many pelagic species, this surface ocean feature is no barrier to the benthos[2]. Indeed, the northward movement of deep water formed in the Weddell Sea makes close faunal connections between the Southern Ocean and other ocean basins more likely[2,21]. ANDEEP samples yielded many cosmopolitan foraminiferal morphospecies[15]. Molecular genetic studies of three such species support the conspecificity of geographically distant populations[22]. The majority of internal transcribed spacer rDNA sequences from *Epistominella exigua, Cibicidoides wuellerstorfi* and *Oridorsalis umbonatus,* obtained in the Weddell Sea and Arctic Ocean (water depths 572–4,975 m), were almost identical, despite this being one of the most rapidly evolving region of the nuclear genome (Supplementary Table 4). These results provide the first molecular evidence for bipolar and possibly global distributions among deep-sea benthic organisms of meiofaunal size. In the case of *E. exigua,* the species for which we have most information, a study of population genetics suggests substantial levels of gene flow from Antarctic (2,600–4,600 m) to Arctic populations but extremely low levels in the opposite direction. In contrast, foraminiferal species living on the shelf exhibit a much stronger divergence between Arctic and Antarctic populations (Fig. 4; Supplementary Table 5) Among metazoans, more than half of all polychaete species found during the ANDEEP expeditions also occur from north of the Polar Front and about 20% occur north of the equator (Supplementary Fig. 5). The first comprehensive analysis of hexactinellid phylogeny on the basis of molecular data revealed close phylogenetic links between the Arctic and the Antarctic taxa, for example, species of the genus *Caulophacus.* Together with palaeontological data, this pattern indicates an ancient colonization of the Southern Ocean deep sea by glass sponges, possibly dating from the Early Cenozoic after this group disappeared from the shelf habitats and migrated into the deep sea at the end of their Cretaceous radiation.

Other evidence, however, points to some strong contrasts between deep Southern Ocean faunas and those from other oceans. The isopods in ANDEEP samples are particularly distinctive with many species presently unknown outside the Southern Ocean. Among the Munnopsidae, the most successful deep-sea family, as well as in other important Southern Ocean taxa (for example, Desmosomatidae, Haploniscidae and Ischnomesidae), >95% of the ANDEEP species are undescribed. Although we know that some species complexes have radiated in the deep Southern Ocean (for example, the Haploniscidae), it is unclear whether they have evolved here and subsequently spread into other ocean basins. The few Southern Ocean deep-sea isopod species that have been described show closest zoogeographical links to Atlantic faunas. Many ANDEEP ostracod species recognized on the basis of their soft-part anatomy are presently unknown outside the Southern Ocean. Furthermore, the markers 16S rDNA, cytochrome oxidase I mtDNA and internal transcribed spacer rDNA indicate large genetic distances between intraspecific populations collected in the western and eastern Weddell Sea. The ostracod family Macrocypprididae was common in ANDEEP material, but usually rare in deep-sea samples from other oceans. The majority (~75%) of mollusc species are currently unknown outside the Southern Ocean, and wide-ranging Atlantic deep-sea species, such as the gastropod *Benthonella tenella,* were not collected. Certain polychaete families (Spionidae, Paraonidae, and Cirratulidae), which are common and species-rich at temperate latitudes, were rare and represented by few species[11]. Many Southern Ocean nematode species are unrecorded and apparently confined to particular parts of the Weddell Sea, although some have wider distributions.

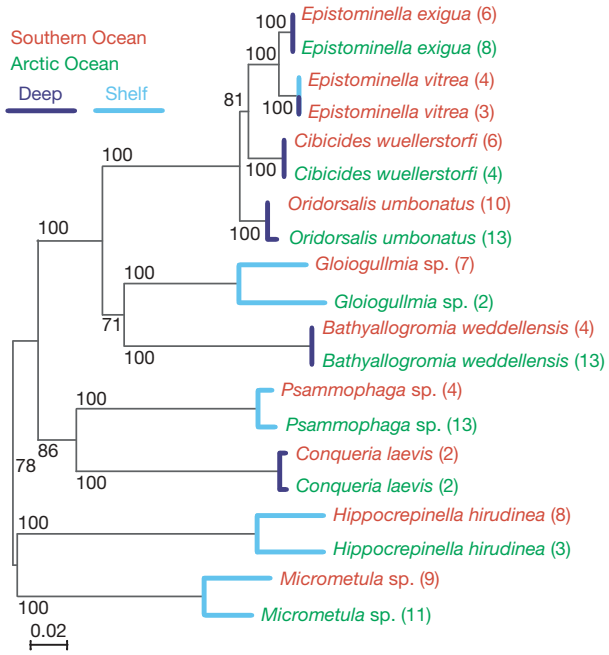


Figure 4 | Phylogenetic relationships of deep-sea foraminifera. Phylogenetic relations between Southern Ocean and Arctic Ocean populations of selected shelf and deep-sea foraminiferan species on the basis of small subunit rDNA sequences. All the shelf species examined are highly divergent, whereas most of the deep-sea species have almost identical sequences. The genetically homogenous species *Epistominella vitrea,* which occurs at shallow and deep sites in the Southern Ocean, is also illustrated. The number of obtained sequences is indicated after species names. Detailed data on sampling sites and sequence divergence are given in Supplementary Material (Supplementary Table 4).

These biogeographic patterns may be linked to larval ecology. The isopods, ostracods and nematodes have poor dispersal capabilities (isopods, for example, are brooders[10]) and hence reduced gene flow, making restricted species distributions more plausible, albeit difficult to prove. The recent discovery of cryptic isopod species in the Southern Ocean and Angola Basin support this assumption[23]. On the other hand, groups in which species are widely distributed, for example, Foraminifera, have easily transported propagules or larval stages[24].

Our data can be used to test the suggested existence of global latitudinal diversity trends in the deep sea[3,5,25,26]. A decrease in isopod biodiversity with increasing latitude in the Southern Hemisphere has been claimed on the basis of data from up to 40° S (ref. 26), although species richness data point to considerable regional variation for isopods on continental slopes[3,25–27]. ANDEEP data, however, indicate high levels of diversity among certain animal groups in the deep Southern Ocean. Values of $E(S_{100})$ (expected number of species per 100 individuals) for isopods from the Weddell and Scotia Seas (22–25 species) are comparable to those from many temperate and tropical slopes in the Southern Hemisphere[25]. Reported $E(S_{100})$ and $E(S_{200})$ values for bivalves, gastropods and isopods from the southeastern Weddell Sea (500–2,000 m) fall within the upper range of those from the tropics[27]. Isopod diversity is even higher ($E(S_{100}) = 55$ species at 3,100 m) than on the southern Australian shelf (40 species)[10] and the Argentine basin[28,29]. Our results further indicate that these high levels of biodiversity extend to abyssal depths; for example, $E(S_{100})$ for isopods is >30 at almost 5,000 m (Supplementary Fig. 4). Possibly, this is a regional phenomenon and diversity is lower in other parts of the deep Southern Ocean. Nevertheless, the extraordinary richness of species in ANDEEP samples when compared with literature data from tropical and temperate sites in the South Atlantic challenges the idea that deep-sea diversity is depressed towards the Antarctic continent. Moreover, the Atlantic is the youngest ocean. A longer period of faunal evolution and diversification in the Pacific and Indian Oceans may have led to greater species richness than in the Atlantic sector of the Southern Ocean. Continuing studies at the morphological and molecular levels of bathyal and abyssal Antarctic faunas may provide further insights into the evolutionary processes that have helped to shape these important communities.

METHODS

A sediment profile imaging system (SPI) was used to obtain photographs and video footage of the seafloor and sediment profiles. Two Barnett-type, hydraulically damped multicorers equipped with 57-mm and 100-mm diameter tubes (25.5 and 78.6 cm^2 surface area respectively), and two box corers (2,500 cm^2 surface area) subdivided into 100 cm^2 'vegematic' subcores recovered sediment samples with well-defined surface areas. Multicores were used principally for the meiofauna (32–300 μm size range) and characterization of sediment parameters; box corers were used to sample the macrofauna (>300 μm). Two towed gears, an epibenthic sledge for macrofauna and an Agassiz trawl for megafauna, obtained material from a much larger but less well-defined area of seafloor and provided most of the taxonomic material in these two size classes. The sledge comprised a lower epibenthic net (500 μm mesh size) and an upper suprabenthic net (300 μm mesh size), each 1 m wide. It was pulled across the seabed for 10 min at a mean velocity of 1 knot. Total distances covered varied from 711 to 6,464 m. The 3-m-wide Agassiz trawl had a cod-end mesh size of 500 μm, except at stations 74, 78 and 81, at which a 10-mm mesh was used. It was fished in a similar way to the epibenthic sledge. Sledge catches were usually fairly clean, whereas the trawl often recovered large quantities of mud which required sieving. Specimens for morphological study were sorted either on board or in a shore laboratory. Metazoans for DNA extraction were picked from samples that had been transferred immediately after collection into pre-cooled 96% ethanol and kept for at least 48 h at −20 °C. Foraminifera for molecular study were picked from unfixed samples in an ice-chilled dish and either frozen in liquid nitrogen or the DNA individually isolated using guanidine lysis buffer.

Received 13 September 2006; accepted 10 April 2007.

1. Clarke, A. & Johnston, N. M. Antarctic marine benthic diversity. *Oceanogr. Mar. Biol.* **41,** 47–114 (2003).
2. Clarke, A. in *Ecosystems of the World. Ecosystems of the Deep Oceans*, Vol. 28 (ed. Tyler, P. A.) 239–260 (Amsterdam, Elsevier, 2003).
3. Gage, J. D. Diversity in deep-sea benthic Macrofauna: the importance of local ecology, the larger scale, history and the Antarctic. *Deep-sea Res. II* **51,** 1689–1708 (2004).
4. Brandt, A. *et al.* The biodiversity of the deep Southern Ocean benthos. *Phil. Trans. Roy. Soc. B* **362,** 39–66 (2007).
5. Rex, M. A., Crame, A., Stuart, C. T. & Clarke, A. Large-scale biogeographic patterns in marine mollusks: a confluence of history and productivity. *Ecology* **86,** 2288–2297 (2005).
6. Sanders, H. L. & Hessler, R. R. Ecology of the deep-sea benthos. *Science* **163,** 1419–1424 (1969).
7. McCann, K. S. The diversity–stability debate. *Nature* **405,** 228–233 (2000).
8. Snelgrove, P. V. R. & Smith, C. R. A riot of species in an environmental calm: the paradox of the species-rich deep sea floor. *Oceanogr. Mar. Biol.* **40,** 311–342 (2002).
9. Levin, L. A. *et al.* Environmental influences on regional deep-sea species diversity. *Annu. Rev. Ecol. Syst.* **32,** 51–93 (2001).
10. Brandt, A. On the origin and evolution of Antarctic Peracarida (Crustacea, Malacostraca). *Sci. Mar.* **63,** 261–274 (1999).
11. Hilbig, B. & Blake, J. A. Deep-sea polychaete communities in the northeast Pacific Ocean off the Gulf of the Farallones, California. *Bull. Mar. Sci.* **78,** 243–269 (2006).
12. Brey, T. *et al.* Do Antarctic benthic invertebrates show an extended level of eurybathy? *Antarct. Sci.* **8,** 3–6 (1996).
13. Tyler, P. A., Young, C. M. & Clarke, A. Temperature and pressure tolerances of embryos and larvae of the Antarctic sea urchin *Sterechinus neumayeri* (Echinodermata: Echinoidea): potential for deep-sea invasion from high latitudes. *Mar. Ecol. Prog. Ser.* **192,** 173–180 (2000).
14. Pawlowski, J. *et al.* A note on the genetic similarity between shallow- and deep-water specimens of the Antarctic foraminifer *Epistominella vitrea. Deep-sea Res. II* (submitted).
15. Cornelius, N. & Gooday, A. J. 'Live' (stained) deep-sea benthic foraminiferans in the western Weddell Sea: trends in abundance, diversity and taxonomic composition along a depth transect. *Deep-sea Res. II* **51,** 1571–1603 (2004).
16. Murray, J. W. & Pudsey, C. J. Living (stained) and dead foraminifera from the newly ice-free Larsen Ice Shelf, Weddell Sea, Antarctica: ecology and taphonomy. *Mar. Micropaleontol.* **53,** 67–81 (2004).
17. De Mesel, I., Lee, H. J., Vanhove, S., Vincx, M. & Vanreusel, A. Species diversity and distribution within the deep-sea nematode genus *Acantholaimus* on the continental shelf and slope in Antarctica. *Polar Biol.* **29,** 860–871 (2006).
18. Hessler, R. R. & Thistle, D. On the place of origin of deep-sea isopods. *Mar. Biol.* **32,** 155–165 (1975).
19. Kussakin, G. O. Peculiarities of the geographical and vertical distribution of marine isopods and the problem of deep sea fauna origin. *Mar. Biol.* **23,** 19–34 (1973).
20. Raupach, M. J., Held, C. & Wägele, J.-W. Multiple colonization of the deep-sea by the Asellota (Crustacea: Peracarida: Asellota). *Deep-sea Res. II* **51,** 1787–1795 (2004).
21. Vinogradova, N. G. Zoogeography of the abyssal and hadal zones. *Adv. Mar. Biol.* **32,** 325–387 (1997).
22. Pawlowski, J. *et al.* Bipolar gene flow in deep-sea benthic foraminifera. *Molecular Ecol.* (submitted).
23. Raupach, M. J., Malyutina, M., Brandt, A. & Wägele, J. W. Molecular data reveal a highly diverse species flock within the munnopsoid deep-sea isopod *Betamorpha fusiformis* (Barnard, 1920) (Crustacea: Isopoda: Asellota) in the Southern Ocean. *Deep-sea Res. II* (submitted).
24. Alve, E. & Goldstein, S. T. Propagule transport as a key method of dispersal in benthic foraminifera (Protista). *Limnol. Oceanogr.* **48,** 2163–2170 (2003).
25. Poore, G. C. B. & Wilson, G. D. F. Marine species richness. *Nature* **361,** 597–598 (1993).
26. Rex, M. A. Stuart, C. T., Hessler, R. R., Allen, J. A., Sanders, H. L., & Wilson, G.D.F. Global-scale latitudinal patterns of species diversity in the deep-sea benthos. *Nature* **365,** 636–639 (1993).
27. Brey, T. *et al.* Antarctic benthic diversity. *Nature* **368,** 297 (1994).
28. Hessler, R. R. & Wilson, G. D. F. in *Evolution, Time and Space: The Emergence of the Biosphere.* (eds Sims, R.W., Price, J.H. & Whalley, P.E.S.) 227–254 (Academic Press, London, 1983).
29. Brandt, A. Zur Besiedlungsgeschichte des antarktischen Schelfes am Beispiel der Isopoda (Crustacea, Malacostraca). *Ber. Polarforsch.* **98,** 1–240 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Financial support for the ANDEEP I–III expeditions was provided by the German Science Foundation. We are also grateful for support from the Swiss National Science Foundation and UK Natural Environment Research Council. S.N.B. received financial support from the DAAD, CAPES, Scar and Synthesys. We thank L. Excoffier, J. Fahrni, J. Guiard, D. Longet and B. Lecroq for help in obtaining the foraminiferal sequence data. We are grateful to D. Fütterer, chief scientist on *Polarstern* cruise ANT XIX/3-4, and E. Fahrbach, chief scientist on *Polarstern* cruise ANT XXII/3, and to the captain and crew of RV *Polarstern*, for help

on board. The DZMB is thanked for financial support of sorting animals from samples and for a research grant for F. Kavanagh, who kindly identified the ANDEEP III Ischnomesidae (Isopoda) and compared these with those species from ANDEEP I and II. We are grateful to many colleagues for constant help on board, helpful and stimulating discussions, and comments, which improved the manuscript. This is ANDEEP publication no. 85 and CAML publication no. 6. This publication also contributes to the CoML field project CeDAMar.

Author Contributions A.B. and A.J.G. assembled and wrote most of the text and were responsible for the theory. J.H. contributed to sedimentology and B. Diaz to sediment analysis, bioturbation and megafauna visible in underwater images. B. Danis and D.C.G. worked on archaebacteria. Foraminiferal data are from N.C., T.C., A.J.G. and J.P.; J.P. was responsible for the molecular work on the Foraminifera. I.D.M. and A.V. delivered the data for Nematoda. A.B., S.B., W.B., M.C., S.K. and M.M. were responsible for the isopod analysis, M.R. was mainly responsible for the molecular isopod data. S.N.B. was responsible for the Ostracoda, B.E. for the Polychaetes, K.L. for the Mollusca and D.J. for the Porifera analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.B. (abrandt@zoologie.uni-hamburg.de).

Vol 447|17 May 2007|doi:10.1038/nature05770

LETTERS

Bypassing genomic imprinting allows seed development

Moritz K. Nowack[1], Reza Shirzadi[2], Nico Dissmeyer[1], Andreas Dolf[3], Elmar Endl[3], Paul E. Grini[2] & Arp Schnittger[1]

In developing progeny of mammals the two parental genomes are differentially expressed according to imprinting marks, and embryos with only a uniparental genetic contribution die[1–3]. Gene expression that is dependent on the parent of origin has also been observed in the offspring of flowering plants, and mutations in the imprinting machinery lead to embryonic lethality, primarily affecting the development of the endosperm—a structure in the seed that nourishes the embryo, analogous to the function of the mammalian placenta[4]. Here we have generated *Arabidopsis thaliana* seeds in which the endosperm is of uniparental, that is, maternal, origin. We demonstrate that imprinting in developing seeds can be bypassed and viable albeit smaller seedlings can develop from seeds lacking a paternal contribution to the endosperm. Bypassing is only possible if the mother is mutant for any of the *FIS*-class genes, which encode Polycomb group chromatin-modifying factors. Thus, these data provide functional evidence that the action of the FIS complex balances the contribution of the paternal genome. As flowering plants have evolved a special reproduction system with a parallel fusion of two female with two male gametes, our findings support the hypothesis that only with the evolution of double fertilization did the action of the *FIS* genes become a requirement for seed development. Furthermore, our data argue for a gametophytic origin of endosperm in flowering plants, thereby supporting a hypothesis raised in 1900 by Eduard Strasburger.

Flowering plants (angiosperms) have evolved to be one of the predominant life forms on earth with more than 250,000 extant species. An important feature, and probably one of the main reasons for this evolutionary success, is the development of an embryo along with a second fertilization product, the endosperm. The endosperm is usually triploid because a homo-diploid female central cell fuses with one of the male gametes. The exact genome dosage of typically two maternal and one paternal genomes seems to be crucial for endosperm development. Raising the maternal contribution in the endosperm has been found to result in smaller seeds comprising fewer endosperm cells and smaller embryos. In contrast, increasing the paternal input results in larger seeds[5–7]. These results have been interpreted in light of the parental conflict theory (kinship theory) according to which mothers and fathers have a different interest in allocation of resources to their offspring[8,9]. Hence, it has been proposed that paternal genes promote seed growth, whereas maternal genes rather reduce growth; or, conversely, that in the maternal genome growth promoting factors are inactivated.

Indeed, flowering plants have been found to imprint certain genes during endosperm development, which is similar to the situation observed for the placenta of mammals[10–14]. Interestingly, the imprinting machinery seems to regulate itself[15–17]. Imprinting involves the recognition of methylated DNA and histone modifications, leading to chromatin rearrangements. In flowering plants, the FIS complex has been implicated in mediating imprinting, and mutants for individual components—for example, *medea* (*mea*), *fertilization-independent seed 2* (*fis2*), *fertilization-independent endosperm* (*fie*)—show an over-proliferation phenotype of the endosperm, resulting in embryo lethality later in seed development[18–21]. The molecular nature of this mutant phenotype, however, has not been resolved, and a growth factor that is specifically expressed during seed development has not been found. Also, although isolated gametes could be fused *in vitro*, attempts to dissect double fertilization and imprinting at the organ level have not yet been successful[22]. Thus, the actual role of imprinting in seed development remains obscure.

Previously we have characterized a mutant for the *Arabidopsis thaliana* $Cdc2^{+}$/*Cdc28* homologue, *CDKA;1*, in which pollen with only one gamete is produced[23]. This mutant pollen causes a paternal effect, leading to seed abortion. Because this pollen was found to exclusively fertilize the egg cell and not the central cell, the *cdka;1* mutant offers a unique possibility to study the action of the paternal genome and the importance of imprinting during seed development.

To trace the fate of plants fertilized with *cdka;1* mutant pollen, we generated a marker line by fusing the *CDKA;1* complementary DNA to the yellow fluorescent protein gene (*YFP*) under the control of the *CDKA;1* promotor. This construct could completely rescue the *cdka;1* mutant phenotype (Supplementary Fig. 1a, b). Next, we selected plants that were homozygous *cdka;1* mutant at the endogenous locus and contained only one allele of the *CDKA;1–YFP* rescue construct; hereafter referred to as *cdka;1–yfp*$^{+/-}$. These plants mimicked heterozygous *cdka;1* mutant plants with half of the pollen showing a mutant phenotype (Supplementary Fig. 1c, d). In contrast to the wild-type-like pollen and its progeny after fertilization, this mutant pollen and its fertilization products are indicated by the absence of YFP fluorescence (Supplementary Fig. 1d, and data not shown). Consistent with previous studies, ovules fertilized with YFP-negative *cdka;1* mutant pollen aborted around 3 days after pollination (d.a.p.) at an early globular embryo stage, surrounded by an underdeveloped endosperm (Fig. 1a–f)[23].

We have previously shown that seed development is coordinated by two independent signals: a positive signal coming from the fertilization of the egg cell and the release of a negative signal regulated by the action of the *FIS*-class genes[23]. Fertilization of only the egg cell with the *cdka;1* mutant pollen generates a positive signal that triggers a few rounds of central cell divisions. To promote further endosperm proliferation, we combined the *cdka;1* mutant with the *mea* mutant, in which the block of central cell proliferation is released due to the lack of FIS–Polycomb group complex function.

In self-fertilized *mea* mutants, the endosperm typically over-proliferates and embryo development arrests at the heart stage (Fig. 1g–i)[18,19,21]. We analysed embryo and endosperm development in

[1]University group at the Max Planck Institute for Plant Breeding Research, Max-Delbrück-Laboratorium, Department of Botany III, University of Cologne, Carl-von-Linné-Weg 10, D-50829 Cologne, Germany. [2]Department of Molecular Biosciences, University of Oslo, PO Box 1041 Blindern, N-0316 Oslo, Norway. [3]Institute of Molecular Medicine and Experimental Immunology, Sigmund-Freud-Straβe 25, D-53105 Bonn, Germany.

mea$^{-/-}$ plants crossed with *cdka;1–yfp*$^{+/-}$ plants. Strikingly, approximately 20% of these seeds (63 out of 329) not only overcame the paternally conferred *cdka;1*-abortion point around 3 d.a.p., but also bypassed the maternally derived *mea* arrest around 6–7 d.a.p. (Fig. 1j–l). This restoration of seed development gave rise to the unforeseen hypothesis that these seeds complete embryogenesis with only a diploid endosperm. Consistently, we found that the majority of seeds arrested with a heart-stage embryo were YFP-positive, whereas developing seeds were YFP-negative (Supplementary Fig. 1g, h).

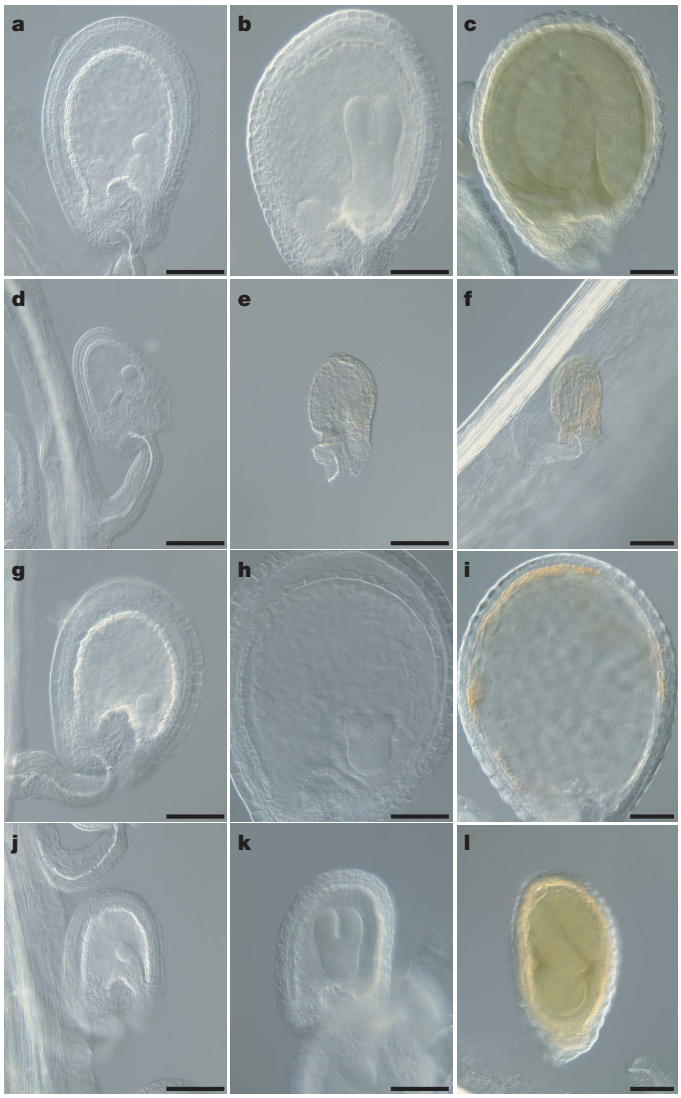


Figure 1 | Rescue of seed abortion in *mea*$^{-/-}$ × *cdka;1–yfp*$^{+/-}$ crosses. Micrographs of cleared seeds. **a–c**, Seed development in L*er* × *CDKA;1–YFP*$^{+/+}$. **a**, Globular-stage embryo. **b**, Torpedo-stage embryo. **c**, Seed nearing maturity. **d–f**, Seed development in L*er* × *cdka;1–yfp*$^{+/-}$ (YFP-negative seeds). **d**, Aborting seed with embryo arrested at early globular stage. The aborted seeds decay (**e**) and (**f**). **g–i**, Seed development in *mea*$^{-/-}$ × *CDKA;1–YFP*$^{+/+}$. **g**, Globular-stage embryo with syncytial endosperm nuclei, similar to wild type. **h**, Seed with an abnormal heart-stage embryo typical for *mea*-conferred seed abortion. **i**, Seed decaying after *mea*-conferred abortion. **j–l**, Seed development in *mea*$^{-/-}$ × *cdka;1–yfp*$^{+/-}$ (YFP-negative seeds). **j**, Globular stage embryo with syncytial endosperm nuclei. **k**, About 25% of the seeds ($n = 112$) are smaller than wild type but embryo development continues. **l**, About 10% of the seeds ($n = 134$) contain small, yet viable-looking embryos. d.a.p., days after pollination. *CDKA;1–YFP*$^{+/+}$ is homozygous for *Pro*$_{CDKA;1}$-*CDKA;1–YFP* in *cdka;1*$^{-/-}$. *cdka;1–yfp*$^{+/-}$ is heterozygous for *Pro*$_{CDKA;1}$*:CDKA;1–YFP* in *cdka;1*$^{-/-}$. YFP-negative, seeds without the *CDKA;1–YFP* transgene. Scale bars, 100 μm.

YFP-negative seeds with a *mea*$^{-/-}$ maternal mutant background showed a characteristic reduction in size in comparison to wild type that became already obvious at 4–5 d.a.p. (Fig. 1j–l; Supplementary Fig. 3a). To examine whether this small seed size phenotype is restricted to the *mea* allele, two other *fis*-class mutants, *fis2* and *fie*, were crossed with *cdka;1*$^{+/-}$ mutants. Similar to *mea* mutants, *fis2* and *fie* have a strong maternal effect and embryos abort around the heart stage[18,20]. However, in the progeny of all crosses we observed a new class of seeds that completed embryogenesis but had an obvious reduction in size similar to that seen for the rescued *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ seeds (data not shown).

We verified that the central cell in the small seeds of the *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ crosses did not receive a paternal genome by performing laser dissection microscopy and subsequent PCR analysis for an accession-specific simple sequence length polymorphism (SSLP) (Fig. 2a–d). In control crosses of *mea*$^{-/-}$ (in the L*er* genetic background) with Col, both maternal and paternal genomes could be detected in both the embryo and the endosperm in approximately 90% of cases (Fig. 2d). In embryos from small *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ seeds, both bands could also be obtained in about 90% of cases. Conversely, in the endosperm of these seeds only the maternal L*er* band could be amplified in 90% of cases (Fig. 2d).

In addition, we measured ploidy levels of developing seeds by flow cytometry to determine whether the endosperm of these small seeds is indeed homoparental diploid rather than biparental triploid. In contrast to the combination of diploid and triploid peaks in wild-type seeds, we observed only diploid nuclei in small *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ seeds (Fig. 2e–f, respectively).

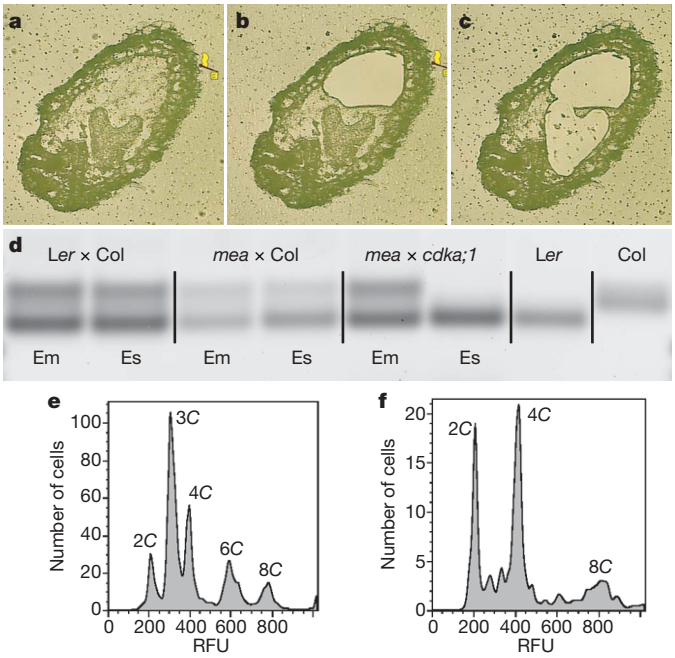


Figure 2 | The small *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ seeds develop with a uniparental diploid endosperm. **a–c**, Micrographs of a *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ rescued-seed section mounted for laser microdissection. **a**, Before dissection. **b**, After collecting endosperm tissue. **c**, After collecting embryo tissue. **d**, Agarose gel showing the products of an SSLP-PCR. The L*er* genome produces a smaller fragment; the Col genome a larger fragment. Ninety per cent of L*er* × Col seeds ($n = 16$) and of *mea*$^{-/-}$ × Col seeds ($n = 19$) have both accession-specific bands in the endosperm. In contrast, 90% of *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ seeds ($n = 16$) contained endosperms only producing a L*er*-band. **e**, **f**, Flow cytometry ploidy analysis of *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ seeds at 6 d.a.p. **e**, DNA profile of wild-type seeds, showing a mixture of diploid (2*C*, 4*C*, 8*C*) and triploid (3*C*, 6*C*) peaks. **f**, DNA profile of small *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ rescued-seeds lacking triploid peaks. L*er*, Landsberg *erecta*; Col, Columbia; *mea*, homozygous *medea* mutant; *cdka;1*, heterozygous *cdka;1* mutant; Em, embryo; Es, endosperm; RFU, relative fluorescence units.

The endosperm in fertilized *fis*-class mutants shows a heterochronic defect; that is, it never reaches maturity[24]. To address the fate of the endosperm in the small seeds originating from the different *fis*-class mutants crossed with *cdka;1*$^{+/-}$, various marker lines were used that are expressed in a development-specific manner. The analysis of these markers revealed that endosperm development in the small *fis*-class-mutant × *cdka;1*$^{+/-}$ seeds is distinct from fertilized *fis*-class mutant seeds and qualitatively resembles wild-type endosperm (Supplementary Fig. 2a–f). A hallmark of endosperm differentiation in *Arabidopsis* is cellularization, which is initiated when the embryo reaches the heart stage (Supplementary Fig. 2g). Whereas cell walls are absent in the endosperm of *mea* mutants (Supplementary Fig. 2h), we found that in small *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ seeds the endosperm was completely cellularized at 6 d.a.p. (Supplementary Fig. 2i). Therefore, the paternal genome does not seem to be required for the adoption of wild-type endosperm qualities, and the *FIS*-class genes are dispensable for the development of a functional endosperm. The diploid endosperm of *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ seeds is genotypically identical to the autonomous endosperm in unfertilized *fis*-class mutant ovules in which the central cell also starts to proliferate. The major difference between these ovules is the presence of a developing embryo. Thus, our data stress a function of the embryo in directing seed growth and suggest repeated cross-talk between embryo and endosperm throughout seed development. First, during very early stages of seed development the zygote stimulates initial central cell divisions as described previously[23]. Here, a second round of embryonic signalling is unravelled because in *fis*-class mutants further differentiation of the diploid endosperm seems to be possible only if an embryo is present.

Except for their reduced size in comparison to wild type, YFP-negative embryos in small *mea*$^{-/-}$ × *cdka;1–yfp*$^{+/-}$ seeds resembled wild-type embryos (see also Fig. 1j–l). *YABBY* messenger RNA, which is a marker for abaxial cell fate, was typically expressed in the two cotyledons at early heart stage (Supplementary Fig. 4a, b). The shoot apical meristem seemed to be properly established, as demonstrated by detection of *WUSCHEL* and *SHOOT MERISTEMLESS* mRNA at a typical region in the axis between the two cotyledons (Supplementary Fig. 4c–f). After completion of embryogenesis, the small seeds that developed with a solely maternal endosperm were found to germinate at frequencies similar to wild-type on agar plates (Supplementary Fig. 3b). However, the size difference of these embryos seemed to be maintained during postembryonic development, and *mea*$^{-/-}$ × *cdka;1–yfp*$^{+/-}$ seedlings were also smaller than controls, as judged by their reduced root lengths after germination (Supplementary Fig. 3c–e). Also, further growth and development of the *mea*$^{-/-}$ × *cdka;1–yfp*$^{+/-}$ plants are delayed, as indicated by smaller rosettes (Supplementary Fig. 3f–l) and a delayed flowering (Supplementary Table 1). Nonetheless, *mea*$^{-/-}$ × *cdka;1–yfp*$^{+/-}$ plants were fertile and could produce viable offspring. At the end of their lifespan, no significant difference in biomass could be detected between *mea*$^{-/-}$ × *cdka;1–yfp*$^{+/-}$ YFP-positive and YFP-negative F_1 plants hatching from small seeds, as judged by shoot dry weight (Supplementary Fig. 3m).

Table 1 | Transmission frequencies of *cdka;1* and the *fis*-class mutant alleles

Genotype of cross	Genotype of viable progeny (%)			Association factor	*n*
	cdka;1$^{+/-}$	*CDKA;1*$^{+/+}$	*fis*-class$^{+/-}$		
Col × *cdka;1–yfp*$^{+/-}$	6	94	NA	NA	412
mea$^{-/-}$ × *cdka;1–yfp*$^{+/-}$	92	8	100	92	186
Col × *cdka;1*$^{+/-}$	10	90	NA	NA	120
mea$^{-/-}$ × *cdka;1*$^{+/-}$	92	8	100	92	142
fis2-4$^{+/-}$ × *cdka;1*$^{+/-}$	39	61	ND	ND	144
fie-11$^{+/-}$ × *cdka;1*$^{+/-}$	36	64	18	100	138

cdka;1–yfp$^{+/-}$ is a homozygous *cdka;1* mutant heterozygously complemented with a $Pro_{CDKA;1}$-*CDKA;1–YFP*$^{+/-}$ transgene; *n*, number of F_1 plants scored; *fis*-class, mutant alleles for *mea*, *fis2* or *fie*; NA, not applicable; ND, not determined; association factor, number of F_1 plants carrying both a *fis*-class mutant and a *cdka;1* mutant allele divided by the total number of plants carrying a *fis*-class mutant allele in per cent.

Embryo development supported by a diploid endosperm is not a rare event in *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ crosses, as conferred by the transmission rate of the *cdka;1* mutant allele. Among the progeny of wild-type plants pollinated with *cdka;1*$^{+/-}$ mutants, only 10% of the siblings carried the mutant allele owing to the paternal effect conferred by the *cdka;1* pollen (Table 1). Strikingly, we found that in *mea*$^{-/-}$ mutants pollinated with *cdka;1*$^{+/-}$ plants the majority of the offspring (92%, $n = 142$) carried the *cdka;1* mutant allele (similar results were obtained using the *cdka;1–yfp*$^{+/-}$ pollen) (Table 1). This fertilization bias was not gene or allele specific because the viable progeny of crosses with other *mea* alleles and other *fis*-class mutants, that is *fis2*$^{+/-}$ and *fie*$^{+/-}$, showed a similar increase in the *cdka;1* allele frequency (Table 1). Notably, a *fie* mutant allele only could be transmitted through the mother in combination with a fertilization by *cdka;1* mutant pollen.

Restoration of the gametophytic maternal effect in *fis*-class mutants by single fertilization of the egg cell suggests that FIS-mediated imprinting counterbalances the activity of the paternal genome. To investigate this hypothesis we analysed the expression levels of the MADS-box transcription factor *PHERES1* (*PHE1*) by quantitative PCR with reverse transcription (qPCR). *PHE1* is one of the few known genes in *Arabidopsis* to be predominantly expressed from the paternal genome in the endosperm and is a target of the FIS complex[25]. Expression data of wild-type control crosses were compared with *mea*$^{-/-}$ × wild-type and *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ crosses (Fig. 3). Consistent with previous findings, *PHE1* expression is strongly upregulated in *mea* mutants[26]. This effect is more pronounced at 4 d.a.p. than at 2 d.a.p. In contrast, *mea*$^{-/-}$ pollinated with *cdka;1*$^{+/-}$ pollen showed a significant reduction in *PHE1* expression when compared to *mea*$^{-/-}$ pollinated with wild-type pollen. Even though previous experiments have shown *phe1* mutants to be indistinguishable from wild-type plants, expression of *PHE1* in antisense orientation could partially rescue the mutant effect of *mea* and elevate the viability of mutant seeds[25,26]. Thus it is conceivable that the rescue of the *mea* mutant phenotype by *cdka;1*$^{+/-}$ pollen is at least partially due to a reconstitution of a more balanced gene dosage of *PHE1* and related genes.

Here we have demonstrated that sexually reproducing plants can be supported by a homoparental diploid endosperm. The finding that viable seeds can develop without a paternal genome contributing to the endosperm demonstrates that imprinting can be bypassed during seed development. On the basis of the monitoring of *PHE1*

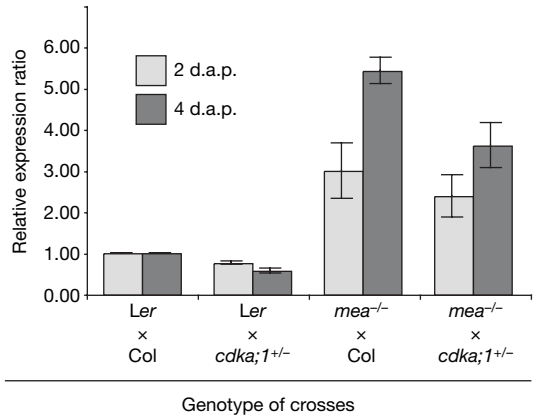


Figure 3 | Quantitative real-time PCR monitoring *PHE1* expression in *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ seeds. Expression levels of *PHE1* were measured by real-time PCR at 2 d.a.p. (light grey bars) and 4 d.a.p. (dark grey bars). A pool of six siliques per biological replica was analysed. The bars represent data from replicate no. 1 (±s.d. of 4 repeats), all biological replicates showed similar results. In the *mea*$^{-/-}$ background, the *PHE1* expression is strongly upregulated, which is most pronounced at 4 d.a.p. On the contrary, in *mea*$^{-/-}$ × *cdka;1*$^{+/-}$ seeds the *PHE1* upregulation at 4 d.a.p. is significantly reduced. For abbreviations, see legends of Fig. 1 and Fig. 2.

expression levels we conclude that the final gene dosage and expression patterns are crucial for seed development: a result consistent with the kinship theory[8,9]. The contribution of the paternal genome and the action of the *FIS*-class genes seem to be opposing mechanisms and might represent sequential layers of regulation that can be stripped off to reveal an ancient core. With this, our findings might also shed light on the evolutionary origin of endosperm.

In the other clade of seed plants, the gymnosperms, offspring are produced only by a single fertilization event. Nonetheless, gymnosperms also support embryo growth with a nourishing tissue of maternal origin, called the prothallium, which is usually well-developed before sperm cell entry[27]. Two hypotheses have been raised to explain the appearance of the endosperm in the angiosperm clade. The first suggests the transformation of a second fertilized embryo into an altruistic nourishing tissue; that is, the endosperm. Thus, endosperm would be of sporophytic origin[28]. Because we have shown here that solely maternally derived endosperm (that is, of gametophytic origin) shows characteristics of wild-type endosperm and is sufficient to sustain complete embryo development, our data provide functional support for the second hypothesis raised by Eduard Strasburger. He proposed, in 1900, that during evolution the female gametophyte was reduced to the central cell of modern angiosperms and that the fusion of the second sperm cell is used as a trigger to start endosperm development[29].

METHODS SUMMARY

Arabidopsis plants used in this study were grown in standard greenhouse conditions. The following genotypes were used: wild-type plants, either the Columbia-0 (Col) or the Landsberg *erecta* (L*er*) accession. Throughout this work the previously characterized *cdka;1-1* allele (SALK_106809.34.90.X) was used[23]. Seeds homozygous for the *mea* allele were obtained from A. Chaudhury and are in the L*er* background[18]. Line KS22 in $fie^{+/-}$ and the enhancer trap GUS line G222 in *mea* were contributed by F. Berger[24]. All genotypes were determined by PCR, by resistance to BASTA or by presence of fluorescent proteins.

Seed morphology was analysed by light microscopy using differential interference contrast optics. Fluorescence in seeds and pollen was analysed with a confocal laser scanning microscope. qPCR was performed on cDNA synthesized from total RNA from siliques. All samples and reference controls were performed in three independent biological replicates and repeated four times.

For a paternity test of the endosperm in *mea* × *cdka;1* seeds, samples of endosperm and embryo tissue were separately collected by laser microdissection and used as templates for SSLP-PCRs discriminating between polymorphisms of the parent accessions L*er* and Col.

For ploidy analysis of the endosperm in *mea* × *cdka;1* seeds, three biological replicates of the nuclei of whole seeds were extracted and analysed with a three laser LSRII analytical flow cytometer.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 November 2006; accepted 23 March 2007.
Published online 29 April 2007.

1. Barton, S. C., Surani, M. A. & Norris, M. L. Role of paternal and maternal genomes in mouse development. *Nature* **311,** 374–376 (1984).
2. Surani, M. A., Barton, S. C. & Norris, M. L. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308,** 548–550 (1984).
3. McGrath, J. & Solter, D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37,** 179–183 (1984).
4. Berger, F. Imprinting—a green variation. *Science* **303,** 483–485 (2004).
5. Scott, R. J., Spielman, M., Bailey, J. & Dickinson, H. G. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* **125,** 3329–3341 (1998).
6. Lin, B.-Y. Association of endosperm reduction with parental imprinting in maize. *Genetics* **100,** 475–486 (1982).
7. Lin, B.-Y. Ploidy barrier to endosperm development in maize. *Genetics* **107,** 103–115 (1984).
8. Haig, D. & Westoby, M. Parent-specific gene expression and the triploid endosperm. *Am. Nat.* **134,** 147–155 (1989).
9. Haig, D. & Westoby, M. Genomic imprinting in endosperm: its effects on seed development in crosses between species and between different ploidies of the same species, and its implications for the evolution of apomixis. *Phil. Trans. R. Soc. Lond. B* **333,** 1–13 (1991).
10. Kohler, C. & Grossniklaus, U. Seed development and genomic imprinting in plants. *Prog. Mol. Subcell. Biol.* **38,** 237–262 (2005).
11. Guitton, A. E. & Berger, F. Control of reproduction by Polycomb group complexes in animals and plants. *Int. J. Dev. Biol.* **49,** 707–716 (2005).
12. Autran, D., Huanca-Mamani, W. & Vielle-Calzada, J. P. Genomic imprinting in plants: the epigenetic version of an Oedipus complex. *Curr. Opin. Plant Biol.* **8,** 19–25 (2005).
13. Gehring, M., Choi, Y. & Fischer, R. L. Imprinting and seed development. *Plant Cell* **16** (Suppl), S203–S213 (2004).
14. Scott, R. J. & Spielman, M. Genomic imprinting in plants and mammals: how life history constrains convergence. *Cytogenet. Genome Res.* **113,** 53–67 (2006).
15. Jullien, P. E., Katz, A., Oliva, M., Ohad, N. & Berger, F. Polycomb group complexes self-regulate imprinting of the Polycomb group gene *MEDEA* in *Arabidopsis*. *Curr. Biol.* **16,** 486–492 (2006).
16. Baroux, C., Gagliardini, V., Page, D. R. & Grossniklaus, U. Dynamic regulatory interactions of Polycomb group genes: *MEDEA* autoregulation is required for imprinted gene expression in *Arabidopsis*. *Genes Dev.* **20,** 1081–1086 (2006).
17. Gehring, M. *et al.* DEMETER DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124,** 495–506 (2006).
18. Chaudhury, A. M. *et al.* Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **94,** 4223–4228 (1997).
19. Grossniklaus, U., Vielle-Calzada, J. P., Hoeppner, M. A. & Gagliano, W. B. Maternal control of embryogenesis by *MEDEA*, a polycomb group gene in *Arabidopsis*. *Science* **280,** 446–450 (1998).
20. Ohad, N. *et al.* A mutation that allows endosperm development without fertilization. *Proc. Natl Acad. Sci. USA* **93,** 5319–5324 (1996).
21. Kiyosue, T. *et al.* Control of fertilization-independent endosperm development by the MEDEA Polycomb gene in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **96,** 4186–4191 (1999).
22. Faure, J. E., Mogensen, H. L., Dumas, C., Lorz, H. & Kranz, E. Karyogamy after electrofusion of single egg and sperm cell protoplasts from maize: cytological evidence and time course. *Plant Cell* **5,** 747–755 (1993).
23. Nowack, M. K. *et al.* A positive signal from the fertilization of the egg cell sets off endosperm proliferation in angiosperm embryogenesis. *Nature Genet.* **38,** 63–67 (2006).
24. Ingouff, M., Haseloff, J. & Berger, F. Polycomb group genes control developmental timing of endosperm. *Plant J.* **42,** 663–674 (2005).
25. Kohler, C., Page, D. R., Gagliardini, V. & Grossniklaus, U. The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. *Nature Genet.* **37,** 28–30 (2005).
26. Kohler, C. *et al.* The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. *Genes Dev.* **17,** 1540–1553 (2003).
27. Maheshwari, P. & Singh, H. The female gametophyte of gymnosperms. *Biological Review* **42,** 88–130 (1967).
28. Sargant, E. Recent work on the results of fertilization in Angiosperms. *Ann. Bot. (Lond.)* **14,** 689–712 (1900).
29. Strasburger, E. Einige Bemerkungen zur Frage nach der "doppelten Befruchtung" bei den Angiospermen. *Botanische Zeitung* **58,** 294–315 (1900).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank F. Berger, E. Erkenbrack, C. Köhler, M. Koornneef and J. Larkin for critical reading and helpful comments on the manuscript. We are grateful to F. Berger for providing mutant and marker lines used in this study. We thank A. Chaudhury for the homozygous *mea (fis1)* mutant line. For the *in situ* hybridization analyses, we thank B. Sæther at NARC, a part of the Norwegian Research Council National Programme for Research in Functional Genomics (FUGE). We thank D. Falkenhahn for embedding and thick-sectioning for the laser dissection microscopy. We are grateful to H. Barbier and M. Reymond for their help with statistical analyses. M.K.N. and N.D. are fellows of the International Max Planck Research School (IMPRS). P.E.G. and R.S. were supported by a grant from the Norwegian Research Council. This work was supported by a grant of the Volkswagen-Stiftung to A.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.S. (schnitt@mpiz-koeln.mpg.de).

METHODS

Plant material and growth conditions. *Arabidopsis* plants used in this study were derived from the Columbia-0 (Col) and the Landsberg *erecta* (L*er*) accessions. Throughout this work the previously characterized *cdka;1-1* allele (SALK_106809.34.90.X) was used[23]. Seeds homozygous for the *mea* allele were obtained from A. Chaudhury and are in the L*er* background[30]. Lines KS117 and KS22 in $fie^{+/-}$, and the enhancer trap GUS line G222 in *mea*, *fis2-4* and *fie-11* were contributed by F. Berger[24]. For genotyping, seeds were germinated on soil or half-strength MS medium and grown under standard green house conditions or in a growth chamber. All genotypes were determined by PCR, by resistance to BASTA or by presence of YFP.

DNA and RNA work. All primers and probes are listed in Supplementary Table 2. For the rescue construct a region of 2,000 base pairs 5′ upstream of the *CDKA;1* start codon together with the *CDKA;1* cDNA fused to a yellow fluorescent protein (YFP) was used.

Total RNA was isolated from siliques using the RNeasy Plant Minikit (QIAGEN) and treated with DNase (TaKaRa) according to the manufacturer's protocol. The RNA concentration was measured twice using a Nanodrop ND-1000 instrument and 3.5 μg of total RNA was used to synthesize cDNA by reverse transcription using Superscript III (Invitrogen). After RNase H treatment at 37 °C for 20 min, a 1:1 dilution of the synthesized cDNA was used in quantitative real-time PCR (qPCR).

qPCR was performed on a Light-cycler LC480 instrument (Roche) according to the manufacturer's protocol. To ensure that the primer combinations did not produce any undesired PCR fragments or primer dimers, a SYBER-GREEN qPCR with melting point analysis was performed using the LightCycler 480 SYBR Green I Master Kit (Roche). Probe-based qPCR with these primers was performed using Universal Probe Library (UPL) hydrolysis probes (Roche) UPL probe no. 147, catalogue no. 04694333001 (*PHE1*) and UPL probe no. 102, catalogue no. 04692209001 (*ACTIN2*) and the LightCycler 480 Probes Master Kit (Roche). All samples and reference controls were performed in three independent biological replicates and repeated four times. The qPCR efficiency was determined independently in all replicates and duplicates by series of tenfold dilutions for each experiment. Calculation of relative expression ratios was performed according to a model previously described[31].

Histology. Pistils and siliques of different developmental stages were prepared as described previously[32]. Dissected siliques were fixed and mounted on microscope slides in a chloral hydrate clearing solution. Light microscopy was performed with a Zeiss Axiophot microscope using DIC optics.

Whole-mount GUS assays were performed as previously described[33] followed by chloral hydrate clearing as described above.

YFP fluorescence in seeds and pollen was analysed with a Leica TCS SP2 AOBS confocal laser scanning microscope. Mature pollen at the stage of anther dehiscence was stained with a 4,6-diamidino-2-phenylindole (DAPI) solution (2.5 μg ml^{-1} DAPI in 50 mM PBS, pH 7.2, with 0.01% Tween20 and 5% DMSO) for one hour.

For the analysis of endosperm cellularization, seeds were fixed and embedded in LR-White plastic resin after a modified protocol[34]; semi-thin sections (0.7 μm) of seeds were prepared with a Reichert Ultracut R microtome and mounted in Canada balsam.

Paternity test of embryo and endosperm. *Arabidopsis* L*er* plants and *mea* mutants (in a L*er* background) were pollinated with pollen from Col plants and *cdka;1* mutants (in a Col background). Six days after pollination, the seeds were fixed in ethanol:acetic acid (3:1) and embedded in Paraplast+ (Kendall) following a standard procedure for plant tissue preparation.

The embedded seeds were dissected using a Rotationsmikrotom 1512 microtome (Leitz). The sections (11 μm thick) were fixed on plastic-coated MembraneSlides (Palm) and stored at 4 °C. Before further handling, the embedding medium was removed using xylole. After rehydration in an ethanol series (100%; 96%; 70%; 50% ethanol in water; finally pure water for 2 min each), the slides were dried and used for laser dissection microscopy. For this, the seed sections were analysed and dissected with a MircoBeam laser dissection microscope (Palm) and fragments of embryos or endosperms of individual seeds were separately collected and stored at −20 °C. Subsequently, the samples were sonicated using an ultrasonic water bath (Bransonic 42 by Branson) for 6× 30 s and applied as template for an accession-specific PCR using a marker (NGA6) to detect simple sequence length polymorphisms between L*er* and Col. The PCR consisted of 55 cycles of product amplification and was performed with LA Taq polymerase (TaKaRa).

Flow cytometry for seed tissue ploidy analysis. For flow cytometry analysis, seeds were crushed with a pistil in 2 ml test tubes with nuclear extraction buffer (CyStain UV-precise kit by Partec). Seeds were identified by their smaller size. All preparations were subsequently filtered through a 30 μm nylon mesh and stained with nuclear staining solution (CyStain UV-precise kit by Partec) containing DAPI. Flow cytometry was performed on a three laser LSRII analytical flow cytometer (BD Biosciences) using the 405 nm solid state laser for excitation and a 440/40 band pass filter for recording of DAPI fluorescence. The ploidy level, represented by the mean peak position in a DAPI fluorescence intensity histogram, was calibrated against the 2*C* nuclear DNA content peak derived from a preparation of young rosette leaves. Doublets were excluded from the analysis by gating on single nuclei in a DAPI-width versus DAPI-area display according to ref. 35. Data were presented using Flowjo analysis software (Tree Star).

30. Chaudhury, A. M. *et al.* Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **94,** 4223–4228 (1997).
31. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29,** e45 (2001).
32. Grini, P. E., Jurgens, G. & Hulskamp, M. Embryo and endosperm development is disrupted in the female gametophytic capulet mutants of *Arabidopsis*. *Genetics* **162,** 1911–1925 (2002).
33. Sessions, A., Weigel, D. & Yanofsky, M. F. The *Arabidopsis thaliana MERISTEM LAYER 1* promoter specifies epidermal expression in meristems and young primordia. *Plant J.* **20,** 259–263 (1999).
34. Lukowitz, W., Mayer, U. & Jurgens, G. Cytokinesis in the *Arabidopsis* embryo involves the syntaxin-related *KNOLLE* gene product. *Cell* **84,** 61–71 (1996).
35. Wersto, R. P. *et al.* Doublet discrimination in DNA cell-cycle analysis. *Cytometry* **46,** 296–306 (2001).

Vol 447|17 May 2007|doi:10.1038/nature05766

LETTERS

Wnt-dependent *de novo* hair follicle regeneration in adult mouse skin after wounding

Mayumi Ito[1], Zaixin Yang[1], Thomas Andl[1], Chunhua Cui[1], Noori Kim[1], Sarah E. Millar[1] & George Cotsarelis[1]

The mammalian hair follicle is a complex 'mini-organ' thought to form only during development[1]; loss of an adult follicle is considered permanent. However, the possibility that hair follicles develop *de novo* following wounding was raised in studies on rabbits[2,3], mice[4] and even humans fifty years ago[5]. Subsequently, these observations were generally discounted because definitive evidence for follicular neogenesis was not presented[6]. Here we show that, after wounding, hair follicles form *de novo* in genetically normal adult mice. The regenerated hair follicles establish a stem cell population, express known molecular markers of follicle differentiation, produce a hair shaft and progress through all stages of the hair follicle cycle. Lineage analysis demonstrated that the nascent follicles arise from epithelial cells outside of the hair follicle stem cell niche, suggesting that epidermal cells in the wound assume a hair follicle stem cell phenotype. Inhibition of Wnt signalling after re-epithelialization completely abrogates this wounding-induced folliculogenesis, whereas overexpression of Wnt ligand in the epidermis increases the number of regenerated hair follicles. These remarkable regenerative capabilities of the adult support the notion that wounding induces an embryonic phenotype in skin, and that this provides a window for manipulation of hair follicle neogenesis by Wnt proteins. These findings suggest treatments for wounds, hair loss and other degenerative skin disorders.

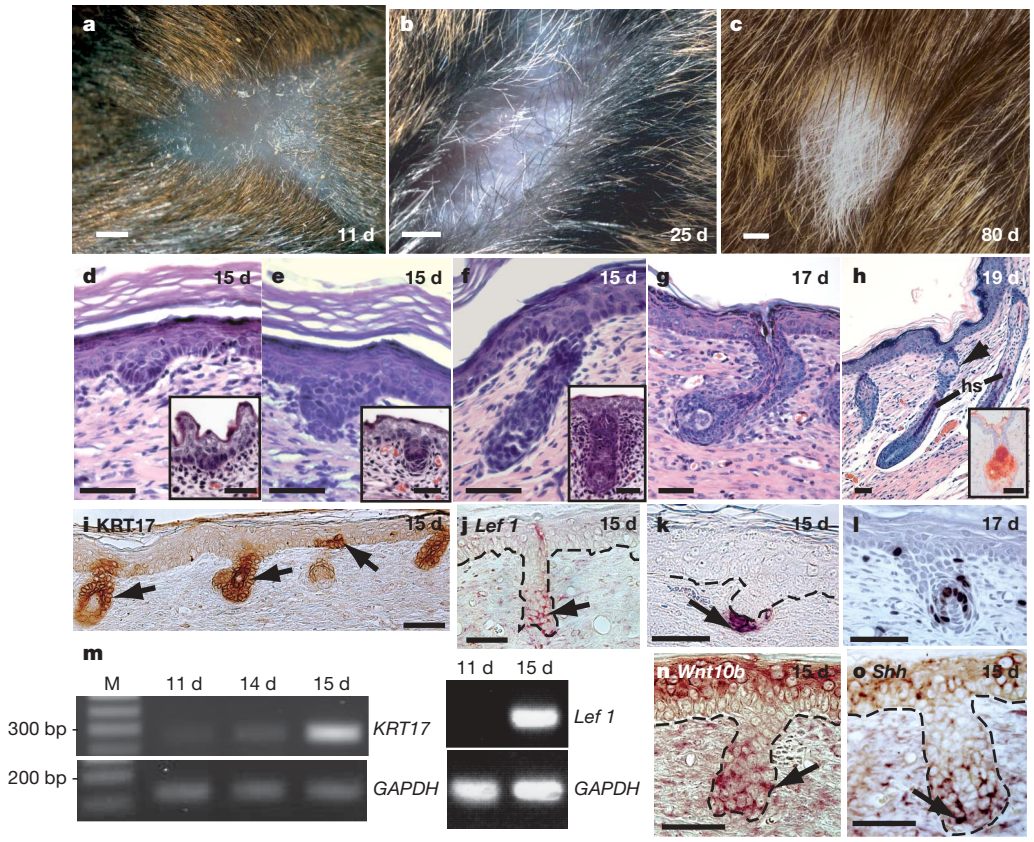


Figure 1 | Hair follicle regeneration recapitulates embryonic development. **a**–**c**, Hair appears in the re-epithelialized wound after several weeks. Scale bar, 1 mm. **d**–**h**, Histology of skin after wounding shows regenerating hair follicles mimicking stages of normal embryonic hair follicle development (examples shown in insets **d**–**f**). Sebaceous gland (arrowhead, **h**) also regenerates from the new follicle (**h**, inset, sebaceous gland stained with oil red O). hs, hair shaft. **i**–**o**, Nascent hair follicles proliferate and express genes and proteins (as indicated) that are associated with hair follicle differentiation, as detected by immunostaining (**i**), BrdU detection (**l**), *in situ* hybridization (**j**, **n**, **o**), alkaline phosphatase activity (**k**) and RT–PCR (**m**). Arrows point to signal. Scale bars, **d**–**l**, **n**, **o**, 50 μm. Time after wounding is in days (d). M, DNA size marker.

[1]Department of Dermatology, Kligman Laboratories, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

During our studies on wound healing in mice, we noticed structures within the centre of large healing wounds that resembled early developing hair follicles. To characterize these structures, we performed timed experiments in which a 1 cm^2 square of full-thickness back skin was excised from 3-week-old mice, at least 2 weeks after the last hair follicles had formed. By 10–11 days after wounding, contracture and re-epithelialization resulted in wound closure and an approximately 0.25 cm^2 area composed of an epidermis and a dermis with no evidence of hair follicles (Fig. 1a). At 14–19 days after wounding, small epidermal downgrowths that resembled developing embryonic hair follicles were present (Fig. 1d–h). Older (7–8-week-old to 10-month-old) mice also showed hair follicle neogenesis after wounding, but larger wounds (2.25 cm^2) were required to trigger follicle formation (Supplementary Table 1). The final rather than initial size of the wound seemed to correlate with hair follicle neogenesis because the larger wound in older mice also yielded a 0.25 cm^2 area immediately after re-epithelialization. (Further characterization of hair follicle neogenesis with respect to density, type and orientation of the hair follicles is presented in Supplementary Table 2, Supplementary Figs 1–3 and Supplementary Notes 1, 2.)

Hair follicles consist of at least ten different epithelial and mesenchymal cell types geared towards the production of hair[7]. In the embryo, hair follicle development begins with the formation of a small cluster of epidermal cells (epithelial placode) that can be detected by expression of cytokeratin 17 (KRT17), an intermediate filament protein[8]. Placodes overlie a dermal condensate, which is identified by alkaline phosphatase activity[9]. Through a series of mesenchymal–epithelial interactions initiated by activation of Wnt and requiring downstream Shh signalling, placode cells proliferate, move downward and engulf the dermal condensate, eventually forming a mature follicle that cyclically produces hair throughout life[10].

We discovered that hair follicle neogenesis following wounding paralleled embryonic follicle development at the molecular level. The regenerated hair follicles expressed KRT17, *Lef1*, alkaline phosphatase, *Wnt10b* and *Shh* (Fig. 1i–o), which is analogous to embryonic follicles[10]. The truly nascent nature of these follicles was demonstrated by absence of expression (by PCR with reverse transcription; RT–PCR) of hair follicle differentiation markers KRT17 (ref. 8) and *Lef1* (refs 10, 11) in the epidermis for several days after wound closure (Fig. lm), and their subsequent appearance coinciding with the development of hair germs and pegs (Fig. li, j, m). The newly formed hair follicles also proliferated normally (Fig. 1l) and generated hair as well as sebaceous glands (Fig. 1c, h).

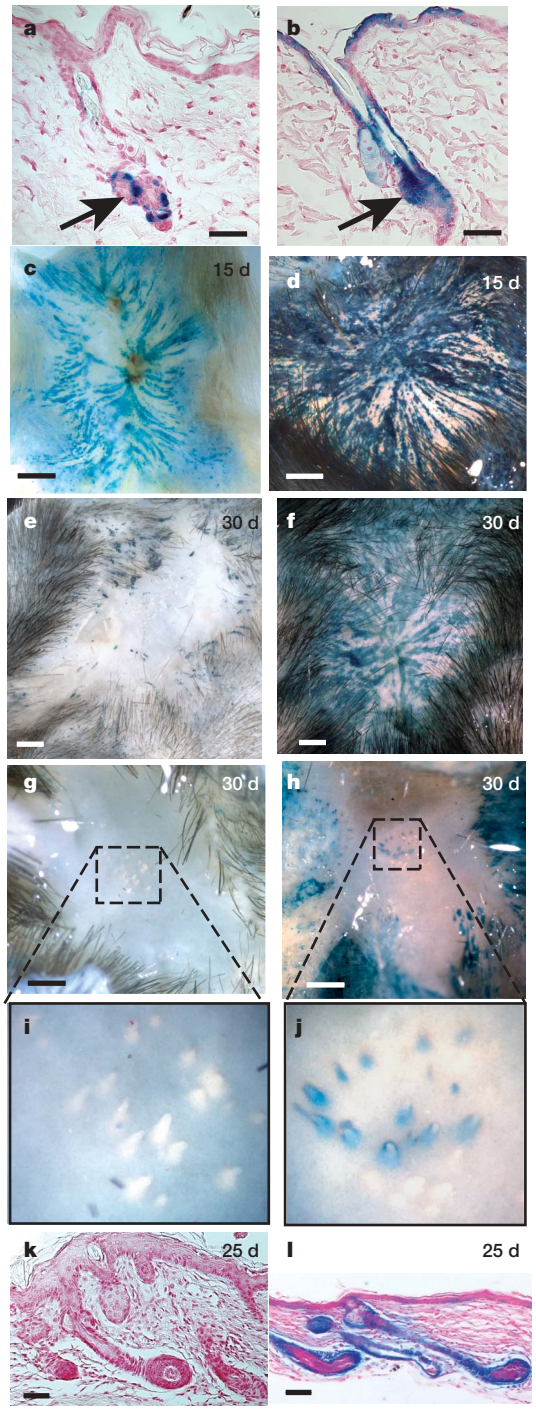


Figure 2 | Regenerated hair follicles originate from non-hair-follicle stem cells. Whole-mounted and sectioned skin, treated to detect *lacZ* expression (blue). **a**, Before wounding, hair follicle stem cells in the bulge (arrow) are labelled by treatment with RU486 in *Tg(Krt1-15-cre/PGR)22Cot;R26R* mice (continued in left panels). **b**, In contrast, both hair follicle stem cells in the bulge (arrow) and ~50% of epidermal cells are labelled in *Krt1-15-CrePR*/R26R* mice (continued in right panels). After wounding, the progeny of the labelled cells move towards the centre of the wound (**c**–**j**), but labelled hair follicles are found predominantly in *Krt1-15-CrePR*/R26R* mice (**g**–**l**). Epidermis was removed in **g**–**j** to visualize the hair follicles (**i**, **j**) of close-ups of **g** and **h**, respectively. Scale bars, 1 mm (**c**–**h**), 50 μm (**a**, **b**, **k**, **l**).

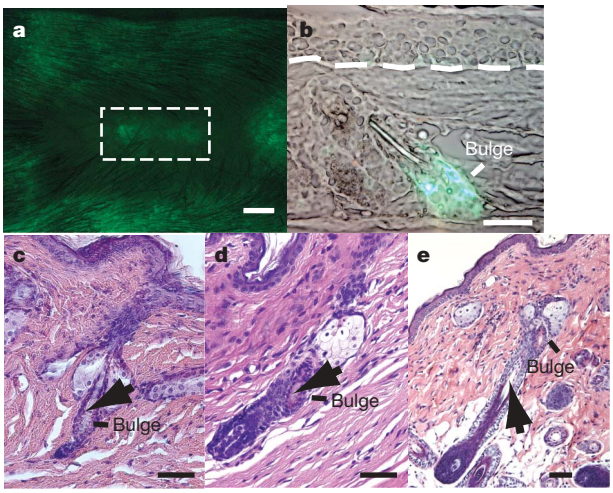


Figure 3 | Functional hair follicle stem cells are re-established in regenerated hair follicles. **a**, **b**, Green fluorescence from the enhanced green fluorescent protein (EGFP) indicates *Krt1-15* promoter activity, a marker of hair follicle stem cells, in *Tg(Krt1-15-EGFP)2Cot* transgenic mouse skin. Thirty-five days after wounding, a cluster of fluorescence is visible in the centre of the wound in whole-mounted skin (**a**). **b**, Histology of this area localizes *Krt1-15* promoter activity to the newly formed bulge in regenerated follicles (dashed line, dermal–epidermal junction). Regenerated hair follicles cycled through stages of rest (35 days, **c**) then growth (40 days, **d**; 45 days, **e**), indicating functional stem cells. (arrows, hair shafts). Scale bars, 1 mm (**a**), 50 μm (**b**–**e**).

We asked next whether the *de novo* follicles arise from hair follicle stem cells in skin bordering the wound. Hair follicle stem cells have been localized to the bulge area[12,13], and bulge cells in follicles surrounding a 4 mm wound send progeny towards the centre of the wound during re-epithelialization[14,15]. However, the majority of bulge cell progeny contribute transiently to the new epidermis and do not persist beyond three weeks[14]. To investigate whether the nascent follicles that developed following larger wounds originated from hair follicle stem cells, we performed genetic lineage analysis using inducible *Tg(Krt1-15-cre/PGR)22Cot;R26R* transgenic mice (Fig. 2, Supplementary Fig. 4 and Supplementary Table 3)[13,14]. These mice express CrePR1, a fusion protein consisting of Cre-recombinase and a truncated progesterone receptor that binds the progesterone antagonist RU486 under the control of the *Krt1-15* promoter, which is active predominantly in bulge cells of adult mouse skin[13,16]. Treatment of adult *Tg(Krt1-15-cre/PGR)22Cot;R26R* mice with RU486 resulted in permanent expression of *lacZ* in bulge cells and in all progeny of labelled bulge cells[13,14]. We discovered that although bulge cell progeny migrated to the centre of the larger wounds, they did not persist. Less than 3% of the new hair follicles were labelled, suggesting that non-hair-follicle bulge cells were the primary source of regenerated follicles (Fig. 2, Supplementary Fig. 4 and Supplementary Table 3).

To investigate further the cells of origin of the regenerated follicles, we examined *Krt1-15-CrePR*;R26R* transgenic mice. These mice possess a mutated progesterone receptor, PR*, containing a larger portion of the progesterone binding domain compared with the PR1 clone[17]. The *Krt1-15-CrePR** mice exhibit Cre recombinase activity during development when the *Krt1-15* promoter is active in the epidermis[16]. In these mice, approximately 70% of bulge and 50% of non-bulge epidermal cells are labelled before wounding (Fig. 2b, Supplemental Table 3; ref. 14). After wounding, approximately half of the regenerated follicles possessed cells expressing *lacZ* (Fig. 2h, j, l). Regenerated follicles were chimaeric for *lacZ* expression (Fig. 2l), indicating that multiple progenitor cells were required for new hair follicle formation, as in development[18]. Thus, together with near absence of *lacZ* expression in regenerated follicles of induced *Tg(Krt1-15-cre/PGR)22Cot;R26R* skin, these data indicate that new follicles originated from cells outside of the hair follicle stem cell niche. The new follicles arose from cells in the epidermis and/or

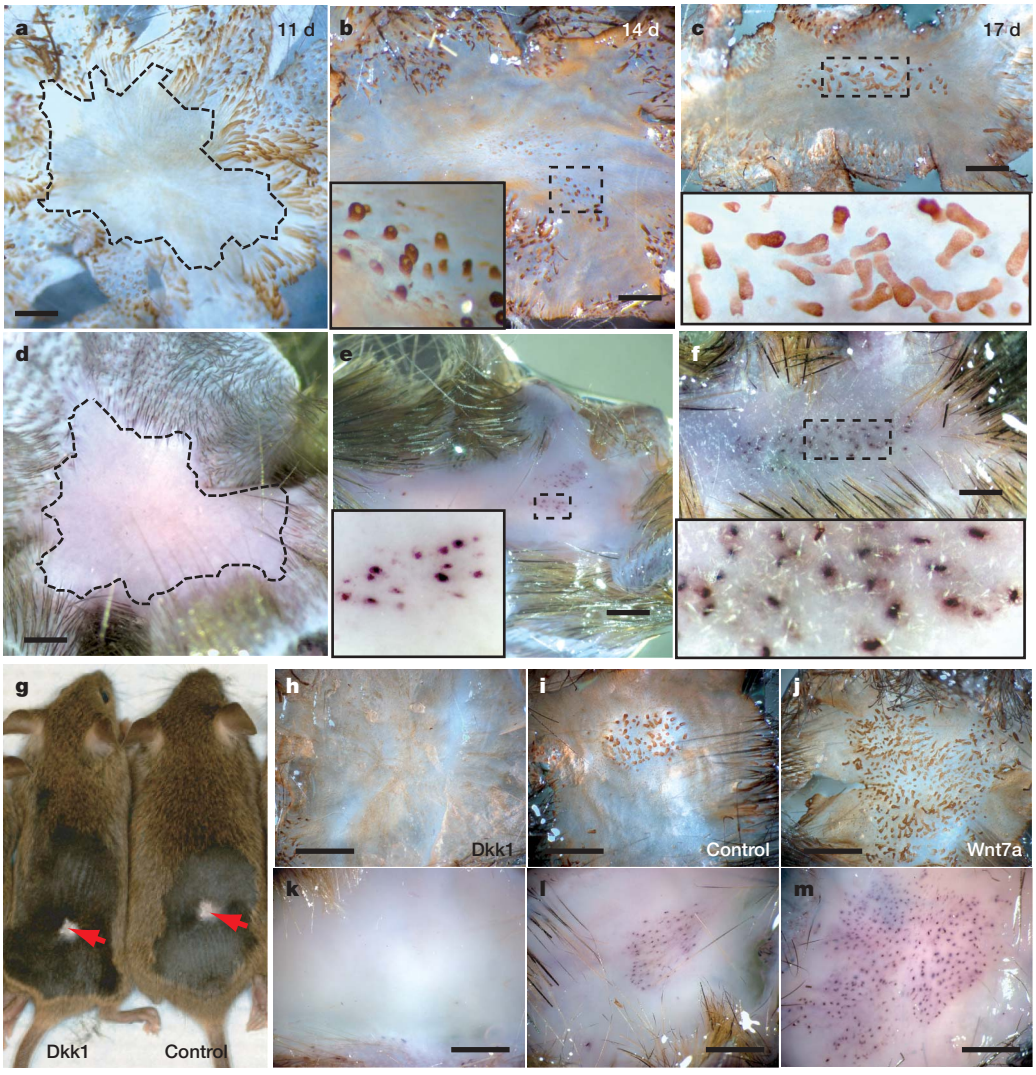


Figure 4 | Hair follicle neogenesis requires Wnt signalling, and overexpression of Wnt potentiates *de novo* follicle formation. Newly regenerating hair follicles in skin stained for KRT17 protein (**a**–**c**, **h**–**j**) and alkaline phosphatase activity (**d**–**f**, **k**–**m**). No hair follicles are visible at 11 days (**a**, **d**), but new follicles form in the centre of the wound at 14 (**b**, **e**) and 17 (**c**, **f**) days post wounding. **a**, **d**, Dashed line marks wound border. **b**, **c**, **e**, **f**, Insets show close-up views of dashed areas of same panel. Dkk1 expression induced in *tetO-Dkk1*;*K5rtTA* mice from 0 to 17 days after wounding does not affect wound closure (**g**), but blocks hair follicle regeneration (**h**, **k**). **j**, **m**, Overexpression of *Wnt7a* in *Krt1-14-Wnt7a* transgenic mice results in increased numbers of hair follicles in a larger area compared with non-transgenic control littermates (**i**, **l**). **h**–**m**, Seventeen days after wounding. Scale bars, 1 mm.

upper portion of the follicle (infundibulum). Both of theses areas are considered to possess stem cells that normally undergo epidermal rather than follicular differentiation[19,20]. Our findings are the first to indicate that non-hair-follicle stem cells in genetically normal adult mice acquire competence to form hair follicles in response to wounding. Whether epidermal or infundibular cells give rise to the new follicles awaits elucidation of reliable markers for these areas. Similarly, identification of the origin of the new hair follicle dermal papilla, which could arise from mesenchymal stem cells, is dependent on a better understanding of skin precursor cells[21].

Because hair follicle stem cells are necessary for follicle survival and cycling[14], we determined whether the regenerated follicles possessed a functional stem cell population. As with normal follicles, we found label-retaining cells and *Krt1-15* promoter activity (Fig. 3, Supplementary Fig. 5), consistent with the re-establishment of a hair follicle stem cell population. The regenerated follicles produced hairs and cycled up to three times within 90 days after wounding (Fig. 3c–e, Supplementary Note 2), indicating the presence of functional stem cells.

The new hairs lacked pigment (Fig. 1c) and associated melanocytes (data not shown), suggesting that the melanocyte stem cell niche was not re-established or that it could not be repopulated[22]. These findings parallel the lack of bulge-derived epithelial cells in the regenerated hair follicles, because, in mice, melanocyte precursors localize to the bulge. Furthermore, melanocytes are not normally present in the interfollicular epidermis of adult mouse back skin[22], precluding repopulation from an epidermal population.

Because Wnt proteins have a key role in normal hair follicle development and cycling[11,23,24], we tested whether Wnt blockade inhibits hair follicle neogenesis following wounding. We induced expression of secreted Dkk1, a Wnt inhibitor, in *tetO-Dkk1*;*K5rtTA* mice at the time of wounding until 17 days later (Fig. 4g–i, k, l). Although the time to re-epithelialization was normal, hair follicles did not form (Supplementary Table 1). Induction of Dkk1 between days 0 and 10 after wounding did not prevent hair follicle neogenesis. However, transient induction of Dkk1 after wound closure inhibited neogenesis and indicated the necessity of Wnt signalling for hair follicle neogenesis (Supplementary Table 1).

We then asked whether overexpression of the secreted ligand, Wnt7a, in *KRT14-Wnt7a* (which targets the epidermis) transgenic mouse epidermis (Supplementary Fig. 6), would enhance hair follicle neogenesis following wounding. Wnt7a has been shown to maintain the hair-follicle-inducing capacity of cultured dermal papilla cells[25]. The overexpression of activated β-catenin, an intracellular Wnt effector, in epidermis induces new hair follicles[26,27], and exogenous Wnt promotes formation of cysts with hair follicle differentiation[28]; however, to date, there has been no evidence that extracellular Wnt ligands can promote actual hair follicle neogenesis in adult skin. Wound closure (time to re-epithelialization) was normal in *KRT14-Wnt7a* mice. However, the transgenic mice developed over twice the number of hair follicles within the wounded area compared with controls (Supplementary Table 1, Fig. 4i, j, l, m). The increased hair follicle number was due to a larger area within the wound that developed follicles ($18 \pm 4\%$ in controls versus $40 \pm 15\%$ in *KRT14-Wnt7a* mice, $P = 0.05$) at the same density as controls (Supplementary Table 2). Thus, excess Wnt in combination with wound healing potentiates regeneration of hair follicles, perhaps by altering cell fate and increasing the number of cells competent to produce hair.

To define better which cell population is responding to Wnt and Dkk during hair follicle regeneration, we examined transgenic mice in which the β-catenin gene *Ctnnb1* had been deleted in the epidermis before wounding using *KRT14-CreER Ctnnb1*$^{flx/flx}$ mice (Supplementary Table 1)[11]. After application of tamoxifen, the *Ctnnb1* gene is deleted in the epidermis of these mice. This resulted in complete inhibition of new hair follicle formation, indicating that Wnt signalling in epidermal keratinocytes is required for hair follicle regeneration (Supplementary Figs 7 and 8).

A major therapeutic goal for those studying stem cells is the ability to coax these cells to differentiate into different cell types with the hope of eventually forming organs through tissue engineering[29]. By taking advantage of existing regenerative programmes, we demonstrated that a wound stimulus is sufficient to trigger regeneration of hair follicles from epithelial cells that do not normally form hair. This *de novo* formation of hair follicles in adult animals recapitulates embryogenesis at the molecular level, and provides a potential window for manipulating the number of hair follicles that form, by exposure to Wnts. This raises the possibility of treating acute wounds with modulators of the Wnt pathway to decrease scar formation, and treating hair loss by regenerating follicles through wounding and Wnt pathway activation.

METHODS SUMMARY

Wounding. C57BL/6J, SJL or mixed strains of C57BL/6J and SJL mice (Jackson laboratory) were anaesthetized with ketamine/xylazine, and a $1\,cm^2$ or $2.25\,cm^2$ cm full thickness excision of skin was made on the mid back of 3×7–8-week-old or 10-month-old mice, respectively. Three-week-old mice were used for all experiments, except as indicated. All animal protocols were approved by the University of Pennsylvania IACUC.

Whole-mount hair follicle neogenesis assay. To detect newly developing hair follicles in the wound, we incubated the skin in 20 mM EDTA in PBS at 37 °C overnight. Then, we gently peeled the epidermal portion off under a Leica MZFLIII dissecting microscope, fixed the epidermis in formalin for 1 h, rinsed with PBS, blocked with 0.6% H_2O_2 in methanol, and performed immunostaining for KRT17 in 1.5 ml eppendorf tubes. We fixed the dermis in acetone at 4 °C overnight, rinsed it with PBS, and then incubated with NBT/BCIP (Roche) for 30 min at 37 °C. We stopped the reaction with 20 mM EDTA in PBS.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 August 2006; accepted 20 March 2007.

1. Schmidt-Ullrich, R. & Paus, R. Molecular principles of hair follicle induction and morphogenesis. *Bioessays* **27,** 247–261 (2005).
2. Billingham, R. E. & Russell, P. S. Incomplete wound contracture and the phenomenon of hair neogenesis in rabbit's skin. *Nature* **177,** 791–792 (1956).
3. Breedis, C. Regeneration of hair follicles and sebaceous glands from epithelium of scars in the rabbit. *Cancer Res.* **14,** 575–579 (1954).
4. Lacassagne, A. & Latarjet, R. Action of methylcholanthrene on certain scars of the skin in mice. *Cancer Res.* **6,** 183–188 (1946).
5. Kligman, A. M. & Strauss, J. S. The formation of vellus hair follicles from human adult epidermis. *J. Invest. Dermatol.* **27,** 19–23 (1956).
6. Straile, W. E. in *Advances in Biology of Skin* Vol. 9 (eds Montagna, W. & Dobson, R.) 369–391 (Pergamon Press, New York, 1967).
7. Cotsarelis, G. Epithelial stem cells: a folliculocentric view. *J. Invest. Dermatol.* **126,** 1459–1468 (2006).
8. McGowan, K. M. & Coulombe, P. A. Onset of keratin 17 expression coincides with the definition of major epithelial lineages during skin development. *J. Cell Biol.* **143,** 469–486 (1998).
9. Paus, R. *et al.* A comprehensive guide for the recognition and classification of distinct stages of hair follicle morphogenesis. *J. Invest. Dermatol.* **113,** 523–532 (1999).
10. Millar, S. E. An ideal society? Neighbors of diverse origins interact to create and maintain complex mini-organs in the skin. *PLoS Biol.* **3,** e372 (2005).
11. Huelsken, J., Vogel, R., Erdmann, B., Cotsarelis, G. & Birchmeier, W. β-Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. *Cell* **105,** 533–545 (2001).
12. Cotsarelis, G., Sun, T. T. & Lavker, R. M. Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* **61,** 1329–1337 (1990).
13. Morris, R. J. *et al.* Capturing and profiling adult hair follicle stem cells. *Nature Biotechnol.* **22,** 411–417 (2004).
14. Ito, M. *et al.* Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nature Med.* **11,** 1351–1354 (2005).
15. Levy, V., Lindon, C., Harfe, B. D. & Morgan, B. A. Distinct stem cell populations regenerate the follicle and interfollicular epidermis. *Dev. Cell* **9,** 855–861 (2005).
16. Liu, Y., Lyle, S., Yang, Z. & Cotsarelis, G. Keratin 15 promoter targets putative epithelial stem cells in the hair follicle bulge. *J. Invest. Dermatol.* **121,** 963–968 (2003).
17. Wunderlich, F. T., Wildner, H., Rajewsky, K. & Edenhofer, F. New variants of inducible Cre recombinase: a novel mutant of Cre–PR fusion protein exhibits

enhanced sensitivity and an expanded range of inducibility. *Nucleic Acids Res.* **29,** E47 (2001).
18. Claudinot, S., Nicolas, M., Oshima, H., Rochat, A. & Barrandon, Y. Long-term renewal of hair follicles from clonogenic multipotent stem cells. *Proc. Natl Acad. Sci. USA* **102,** 14677–14682 (2005).
19. Levy, V., Lindon, C., Zheng, Y., Harfe, B. D. & Morgan, B. A. Epidermal stem cells arise from the hair follicle after wounding. *FASEB J.* advance online publication, doi:10.1096/fj.06–6926com (25 January 2007).
20. Taylor, G., Lehrer, M. S., Jensen, P. J., Sun, T. T. & Lavker, R. M. Involvement of follicular stem cells in forming not only the follicle but also the epidermis. *Cell* **102,** 451–461 (2000).
21. Fernandes, K. J. *et al.* A dermal niche for multipotent adult skin-derived precursor cells. *Nature Cell Biol.* **6,** 1082–1093 (2004).
22. Nishimura, E. K. *et al.* Dominant role of the niche in melanocyte stem-cell fate determination. *Nature* **416,** 854–860 (2002).
23. Andl, T., Reddy, S. T., Gaddapara, T. & Millar, S. E. WNT signals are required for the initiation of hair follicle development. *Dev. Cell* **2,** 643–653 (2002).
24. Van Mater, D., Kolligs, F. T., Dlugosz, A. A. & Fearon, E. R. Transient activation of β-catenin signaling in cutaneous keratinocytes is sufficient to trigger the active growth phase of the hair cycle in mice. *Genes Dev.* **17,** 1219–1224 (2003).
25. Kishimoto, J., Burgeson, R. E. & Morgan, B. A. Wnt signaling maintains the hair-inducing activity of the dermal papilla. *Genes Dev.* **14,** 1181–1185 (2000).
26. Gat, U., DasGupta, R., Degenstein, L. & Fuchs, E. De Novo hair follicle morphogenesis and hair tumors in mice expressing a truncated β-catenin in skin. *Cell* **95,** 605–614 (1998).
27. Lo Celso, C., Prowse, D. M. & Watt, F. M. Transient activation of β-catenin signalling in adult mouse epidermis is sufficient to induce new hair follicles but continuous activation is required to maintain hair follicle tumours. *Development* **131,** 1787–1799 (2004).
28. Fathke, C. *et al.* Wnt signaling induces epithelial differentiation during cutaneous wound healing. *BMC Cell Biol.* **7,** 4 (2006).
29. Atala, A. Tissue engineering, stem cells and cloning: current concepts and changing trends. *Expert Opin. Biol. Ther.* **5,** 879–892 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by US National Institutes of Health and, in part, by a grant from the Pennsylvania Department of Health to G.C. We thank P. Sterling for his comments on the manuscript and L. Ash for preparation of histological sections.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.C. (cotsarel@mail.med.upenn.edu).

METHODS

Immunohistochemistry. Immunohistochemistry using antibodies against KRT17 (1:5,000 dilution, a gift from P. Coulombe), BrdU (1:500; Harlan-Seralab), β-catenin (1:2,000, Sigma) was done as previously described[14,30]. Detection of alkaline phosphatase and β-galactosidase activity was done as described[31]. Immunohistochemical studies were repeated at least three times.

***In situ* hybridization.** *In situ* hybridization for Wnt10b, Shh or Lef1 was done using previously described riboprobes and methods[16,23]. Studies were repeated twice.

RT–PCR. Healed skin was excised at 11, 14 and 15 days after wounding. The epidermis was separated from the dermis by incubation with 20 mM EDTA for 2 h at 37 °C. RNA was isolated from the epidermis using RNeasy (Qiagen) according to the manufacturer's instructions. We used 1 μg of total RNA to synthesize cDNA with a superscript cDNA synthesis kit (Invitrogen). PCR was carried out at 94 °C for 30 s, 55 °C for 1 min, and 72 °C for 1 min, for 30 cycles to detect the expression of KRT17 and Lef1 and GAPDH using described primer sequences[23]. These experiments were repeated three times.

Lineage analyses of hair follicle neogenesis. *Tg(Krt1-15-cre/PGR)22Cot;R26R* and *Krt1-15-CrePR*/R26R* transgenic mice were treated with RU486 for 5 days during the first telogen stage[14] and then wounded. Skin samples were collected at 1 and 15 days after wound closure, and processed for β-galactosidase activity, which appeared blue[14]. The β-galactosidase-expressing cells in the skin were observed before and after peeling off the epidermis after EDTA (20 mM in PBS) treatment at 37 °C. 2–3 mice were used at each time point and experiments were repeated twice.

Detection of hair follicle stem cells in the regenerated follicles. We wounded *Tg(Krt1-15-EGFP)2Cot* mice[13] and then observed the fluorescence of the skin under a dissecting microscope 30 days later. We then prepared frozen sections from the centre of the wound encompassing the fluorescence, fixed the tissue in formalin for 10 min, stained the nuclei with Hoechst dye and detected fluorescence with an upright Leica DM4000B microscope equipped with a Leica DC500 digital camera[14]. For detection of label retaining cells, we injected BrdU (50 mg per g body weight; Sigma) intraperitoneally into mice twice daily for 3 days from 20 days after wounding and then fixed the tissues for BrdU detection at 23 and 40 days after wounding[32]. 2–3 mice were used at each time point and experiments were repeated twice.

***tetO-Dkk1;K5rtTA* transgenic mice.** Production of these mice was previously described[33]. Mice were placed on chow formulated with 1 g kg^{-1} doxycycline (BioServ) for either 0–17, 11–14, 12–15 or 0–10 days after wounding.

***KRT14-Wnt7a* transgenic mice.** A mouse Wnt7a cDNA fragment was generated by RT–PCR from mouse embryonic day E14.5 skin RNA using primers 7aF*Bam*HI cgcggatccgcgaacatrgacccggaaagcgcgg and 7aR*Bam*HI tcgcggatccgcgtcacttgcacgtatacatc (accession number M89801, nucleotides 1–18 and nucleotides 1,032–1,050, respectively). The PCR product was digested with *Bam*HI and cloned into the *Bam*HI site of a *KRT14-hGH* transgenic vector expression cassette[34] resulting in *KRT14-Wnt7a-hGH*. Sequencing revealed three differences to accession number M89801: nucleotides 817 and 818 were altered resulting in a leucine to lysine amino acid substitution. In addition, nucleotide 427 was altered and resulted in a tryptophan to arginine change. These changes are consistent with the Wnt7a sequence NM_009527. The same nucleotide alterations also could be found in clones amplified from the mouse keratinocyte cell line PAM212. For producing transgenic mice a 5 kb *Eco*RI fragment of the *KRT14-Wnt7a-hGH* vector was gel purified and micro-injected into fertilized oocytes (B6SJLF1/J x B6SJLF1/J cross, Jackson Laboratories). Transgenic founder animals were identified by PCR (primers 7aF*Bam*HI and 7aSeqR ccagcttcatgttctcctccagg, nucleotides 573–595 of M89801) and Southern blotting (cDNA used as probe). Detailed phenotypic analysis of *KRT14-Wnt7a* mice will be published elsewhere, but notably the hair follicle dermal papilla of these mice contain ~25–38% more cells compared with normal. It is possible that the larger number of dermal papilla cells contributes to the greater number of hair follicles that form after wounding. Control mice in *KRT14-Wnt7a* experiments refer to littermates not carrying the transgene.

***KRT14CreER;Ctnnb1$^{flx/flx}$*.** To generate these conditional β-catenin knockout mice, *Ctnnb1$^{flx/flx}$* mice (on a mixed background of 129×C57Bl6, a gift from W. Birchmeier) were crossed with *KRT14CreER* mice (on a background of CD1, The Jackson Laboratory). *Ctnnb1$^{flx/+}$*;*KRT14CreER* males were crossed with *Ctnnb1$^{flx/flx}$* females. Tamoxifen (1 mg per 10 g body weight; Sigma) was given via intraperitoneal injection from 9 to 14 days after wounding.

Statistics. Hair follicle numbers are expressed as mean ± s.d. The student's two-tailed *t*-test function in Excel was used to calculate *P* values. Details of statistics from experiments are presented in Supplementary Table 1.

30. Liu, F. *et al.* Wnt–β-catenin signaling initiates taste papilla development. *Nature Genet.* **39,** 106–112 (2006).
31. Handjiski, B. K., Eichmuller, S., Hofmann, U., Czarnetzki, B. M. & Paus, R. Alkaline phosphatase activity and localization during the murine hair cycle. *Br. J. Dermatol.* **131,** 303–310 (1994).
32. Ito, M., Kizawa, K., Hamada, K. & Cotsarelis, G. Hair follicle stem cells in the lower bulge form the secondary germ, a biochemically distinct but functionally equivalent progenitor cell population, at the termination of catagen. *Differentiation* **72,** 548–557 (2004).
33. Chu, E. Y. *et al.* Canonical WNT signaling promotes mammary placode development and is essential for initiation of mammary gland morphogenesis. *Development* **131,** 4819–4829 (2004).
34. Millar, S. E. *et al.* WNT signaling in the control of hair growth and structure. *Dev. Biol.* **207,** 133–149 (1999).

LETTERS

SUMOylation regulates kainate-receptor-mediated synaptic transmission

Stéphane Martin[1], Atsushi Nishimune[1], Jack R. Mellor[1] & Jeremy M. Henley[1]

The small ubiquitin-like modifier protein (SUMO) regulates transcriptional activity and the translocation of proteins across the nuclear membrane[1]. The identification of SUMO substrates outside the nucleus is progressing[2] but little is yet known about the wider cellular role of protein SUMOylation. Here we report that in rat hippocampal neurons multiple SUMOylation targets are present at synapses and we show that the kainate receptor subunit GluR6 is a SUMO substrate. SUMOylation of GluR6 regulates endocytosis of the kainate receptor and modifies synaptic transmission. GluR6 exhibits low levels of SUMOylation under resting conditions and is rapidly SUMOylated in response to a kainate but not an *N*-methyl-D-aspartate (NMDA) treatment. Reducing GluR6 SUMOylation using the SUMO-specific isopeptidase SENP-1 prevents kainate-evoked endocytosis of the kainate receptor. Furthermore, a mutated non-SUMOylatable form of GluR6 is not endocytosed in response to kainate in COS-7 cells. Consistent with this, electrophysiological recordings in hippocampal slices demonstrate that kainate-receptor-mediated excitatory postsynaptic currents are decreased by SUMOylation and enhanced by deSUMOylation. These data reveal a previously unsuspected role for SUMO in the regulation of synaptic function.

Kainate receptors (KARs) have key roles in the regulation of synaptic transmission and neuronal excitability in the mammalian brain. At the presynaptic terminal they can modulate neurotransmitter release, and at the postsynaptic membrane they contribute to fast excitatory synaptic transmission[3]. The GluR6 subunit is expressed throughout the brain but is particularly abundant in the hippocampus, where GluR6 subunits can be present in both pre- and postsynaptic KAR complexes[4].

We identified the specific SUMO-conjugating enzyme Ubc9 and the SUMO ligase PIAS3 (Supplementary Figs 1, 2) as GluR6a interactors in a yeast two-hybrid screen. Of the ionotropic glutamate receptor subunits tested, only GluR6a exhibited strong interaction with both enzymes (Fig. 1a). These interactions were verified by co-immunoprecipitation of Ubc9 and PIAS3 with an anti-GluR6 antibody (Supplementary Fig. 1b, c). Truncation mutagenesis mapped the GluR6 interaction domain (Fig. 1b) to residues 882–895 which, alone among the KAR subunits, contains the consensus SUMOylation motif ΨKXE (ref. 5).

We next probed subcellular fractions with specific anti-SUMO-1 antibodies (Fig. 1c, Supplementary Figs 3, 5). Multiple SUMO-conjugated proteins were detected in all fractions, including high levels in the synaptosomal and postsynaptic density fractions (Fig. 1c). SUMOylation is readily reversible[6], so we assessed the effects of the cysteine isopeptidase inhibitor *N*-ethylmaleimide (NEM)[7]. Detection of SUMOylated proteins was markedly reduced in the absence of NEM (Supplementary Fig. 3a). Furthermore, in brain extracts a distinct higher-molecular-mass GluR6 band was detected in the presence of NEM (Fig. 1d). A specific GluR6 band was also detected in anti-SUMO antibody immunoprecipitates from cultures (Fig. 1e, Supplementary Fig. 5a) further confirming that GluR6 is SUMOylated in hippocampal neurons. As a control, the known SUMO substrate RanGAP1 was also retrieved in SUMO immunoprecipitates (Supplementary Fig. 4a, c). Interestingly, however, SUMOylated surface-expressed GluR6 was not detected in resting cultured neurons (Fig. 1e). Consistent with the yeast two-hybrid data, no corresponding band shift in the presence of NEM was observed for the AMPA receptor subunit GluR1 (Fig. 1e, Supplementary Fig. 4b).

As expected, we found high levels of SUMO immunoreactivity in the nucleus of cultured neurons. SUMO immunoreactivity was also distributed throughout dendrites (Fig. 1f–h), where it displayed a punctate distribution similar to those of GluR6, Ubc9 and PIAS3 (Supplementary Fig. 1d, e) and was present at PSD95-positive synapses (Supplementary Fig. 2a, b). Under basal conditions there were low levels of colocalization of surface-expressed GluR6 and SUMO-1 ($5.3 \pm 0.8\%$) but there was some limited overlap between total GluR6 and SUMO-1 immunoreactivity ($28.6 \pm 1.5\%$; Fig. 1f–h), indicating that most SUMOylated GluR6 is intracellular. The low level of colocalization between surface GluR6 and SUMO-1 is consistent with only a small fraction of substrate being SUMOylated at any given time[6,8].

Application of kainate or NMDA evokes internalization of GluR6 into distinct sorting pathways[9]. We therefore tested the effects of these agonists on GluR6 SUMOylation in immunoprecipitation experiments. Kainate (10 μM) rapidly increased the low basal level of GluR6 SUMOylation (Fig. 2a, b). This was detectable after 1 min (data not shown), increased >2-fold (2.4 ± 0.4, $P < 0.001$) at 3 min and reached maximal conjugation within 10 min (3.4 ± 0.4, $P < 0.001$). Glutamate (1 mM) also had similar effects (Supplementary Fig. 6). Despite strong NMDA-evoked internalization[9] there was no increase in GluR6 SUMOylation (Fig. 2a, b). Neither kainate nor NMDA altered global cellular SUMOylation (Supplementary Fig. 7). Similarly, in immunocytochemical assays colocalization between SUMO-1 and internalized GluR6 was $46.7 \pm 3.0\%$ at 3 min and $54.1 \pm 4.9\%$ at 30 min after adding kainate but only $15.9 \pm 1.8\%$ at 3 min and $16.7 \pm 1.7\%$ at 30 min after adding NMDA (Fig. 2c–h).

To explore the role of SUMOylation in agonist-evoked GluR6 endocytosis we expressed green fluorescent protein (GFP)-tagged SUMO specific isopeptidase SENP-1 (ref. 10) to facilitate deSUMOylation or its inactive point mutant SENP-1 C603S (ref. 11) (Supplementary Fig. 8). SENP-1, but not SENP-1 C603S, prevented kainate-evoked GluR6 internalization but had no effect on NMDA-evoked GluR6 internalization (Fig. 3a, b). Thus, GluR6 SUMOylation is required for kainate-evoked internalization and probably occurs at the plasma membrane. Because we did not detect SUMOylated GluR6 at the surface (Fig. 1e) we reasoned that internalization occurs rapidly after SUMOylation. Consistent with this,

[1]MRC Centre for Synaptic Plasticity, Anatomy Department, University Walk, University of Bristol, Bristol, BS8 1TD, UK.

we detected surface-SUMOylated GluR6 in the presence of sucrose that blocks KAR endocytosis (Supplementary Fig. 9).

We next made glutathione *S*-transferase (GST) fusions of wild-type (WT) and K886R GluR6a carboxy-terminal domains for expression in a bacterial SUMOylation assay[12] to determine whether K886 within the consensus SUMOylation motif on GluR6a is the site of SUMOylation. Wild-type, but not K886R, C-terminal GluR6a was robustly SUMOylated (Fig. 3c). To evaluate the role of K886

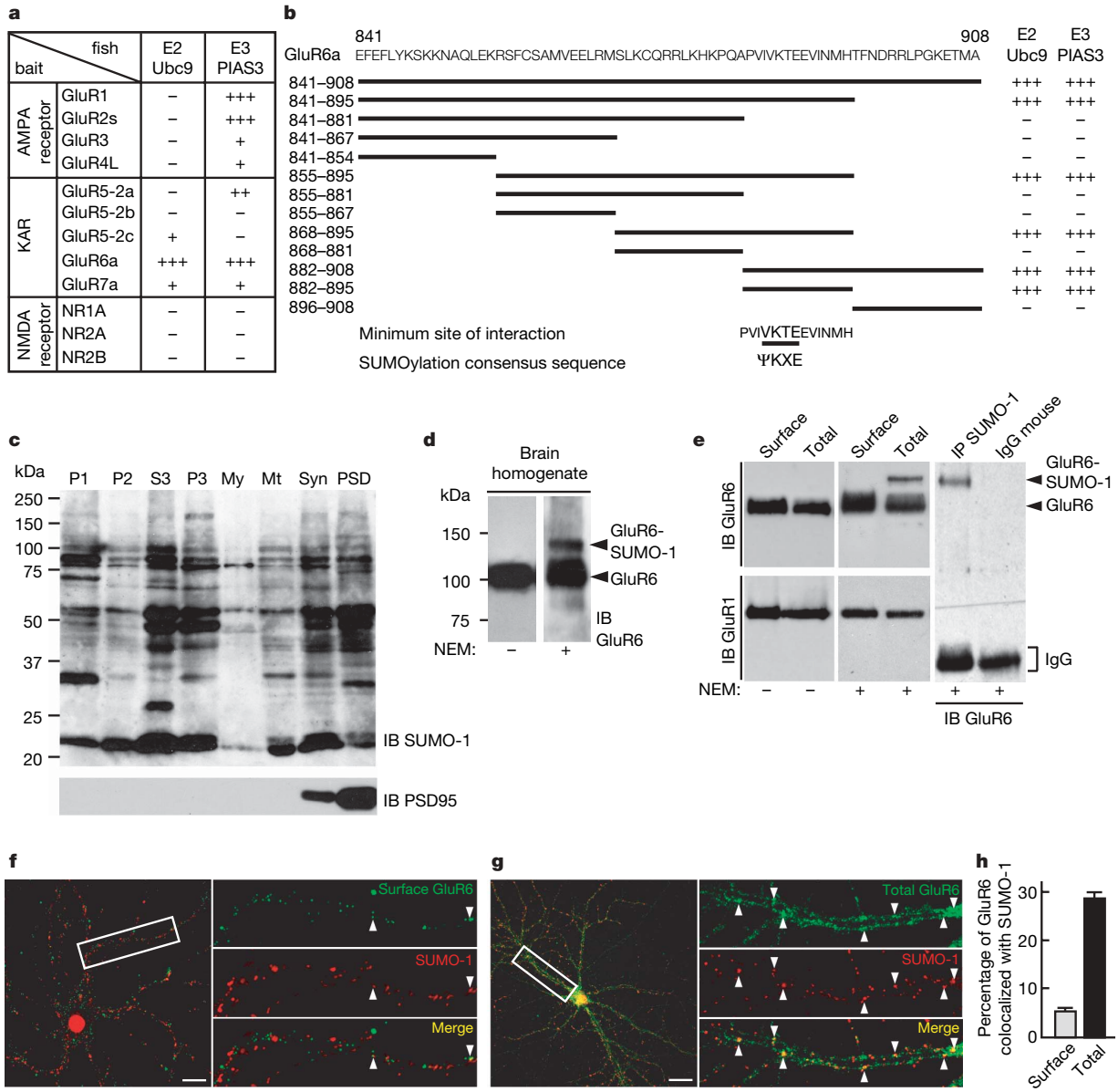


Figure 1 | SUMOylation of GluR6 in brain and cultured hippocampal neurons. a, The SUMO-conjugating enzyme Ubc9 and the SUMO ligase PIAS3 selectively bind to the C-terminal domain of GluR6a in the yeast two-hybrid assay. The C termini of none of the other ionotropic glutamate receptor subunits tested interact strongly with both enzymes, as determined by β-galactosidase reporter activation. **b**, A series of overlapping truncation mutants of the C terminus of GluR6a tested in the yeast two-hybrid assay showed that the domain for both SUMOylation enzymes contains the consensus SUMOylation motif (ΨKXE, where Ψ is the hydrophobic residue). **c**, Immunoblot (IB) analysis of synaptic SUMOylated proteins after rat brain fractionation. Multiple SUMOylated substrates are present in the synaptic fractions. Equal amount of protein (40 μg protein per lane) from each fraction was loaded. P1, crude nuclear fraction; P2, post-nuclear crude membrane fraction; S3, cytosol; P3, microsomes; My, myelin; Mt, mitochondria; Syn, synaptosomes; PSD, post-synaptic density fraction. Immunoblot probed with anti-PSD95 antibody is shown below (2 μg protein per lane). We note that the PSD95 blot is loaded with very low levels of protein to avoid saturation of the signal in the highly enriched synaptic and postsynaptic density fraction lanes. **d**, Immunoblots showing SUMOylated and non-SUMOylated forms of GluR6a in brain homogenate. Note that, in the absence of the cysteine protease inhibitor NEM, no detectable band corresponding to SUMOylated GluR6a was observed, which confirms that the SUMO adduct is highly labile and sensitive to de-conjugation. **e**, Immunoblots showing SUMOylated and non-SUMOylated forms of GluR6a in cultured hippocampal neurons. Immunoprecipitation (IP) experiments using anti-SUMO-1 antibody reveal that a fraction of GluR6a is SUMOylated in cultured hippocampal neurons (right panel). IgG, immunoglobulin. The blots are representative of at least four separate experiments. **f**, **g**, Infrequent but punctate colocalization of surface (**f**) and total GluR6 (**g**) with total SUMO (in green) in unstimulated cultured hippocampal neurons. We note that SUMO is highly concentrated in the nuclear compartment, consistent with its described role in nuclear function, but also clearly show a punctate distribution along the dendrites. Arrowheads denote colocalization. Scale bar, 20 μm. **h**, Quantification of surface and total GluR6 receptor population that colocalize with total SUMO-1 in unstimulated cultured hippocampal neurons as shown in **f** and **g**. Puncta corresponding to GluR6 that colocalized with the SUMO-1 labelling were quantified using the LSM510 version 3.2 software (Zeiss). Analysis was made from at least eight cells for each condition and given as a percentage ± s.e.m.

SUMOylation in KAR endocytosis we compared the properties of the wild-type and K886R-tagged GluR6a tagged with yellow fluorescent protein (YFP). Both wild-type and K886R YFP–GluR6a display similar functional channel properties in COS-7 cells (data not shown). Wild-type YFP–GluR6a was surface-expressed in African green monkey kidney (COS-7) cells with minimal colocalization with SUMO-1 in unstimulated cells (Fig. 3d). Kainate caused a dramatic loss of surface YFP–GluR6a and strong colocalization with SUMO-1 in perimembrane compartments (Fig. 3e). YFP–GluR6a-K886R also trafficked to the plasma membrane (Fig. 3f) but its distribution was unchanged by kainate (Fig. 3f, g). The requirement of K886 for kainate-evoked internalization was further confirmed by surface biotinylation (Fig. 3h, i) and antibody feeding experiments (Supplementary Fig. 10). Kainate evoked massive endocytosis of wild-type YFP–GluR6a (72.9 ± 1.8% at 30 min) but YFP–GluR6a-K886R internalization remained the same as constitutive endocytosis under non-stimulated conditions (12.4 ± 0.6% compared to 11.3 ± 0.9% in the presence of kainate at 30 min).

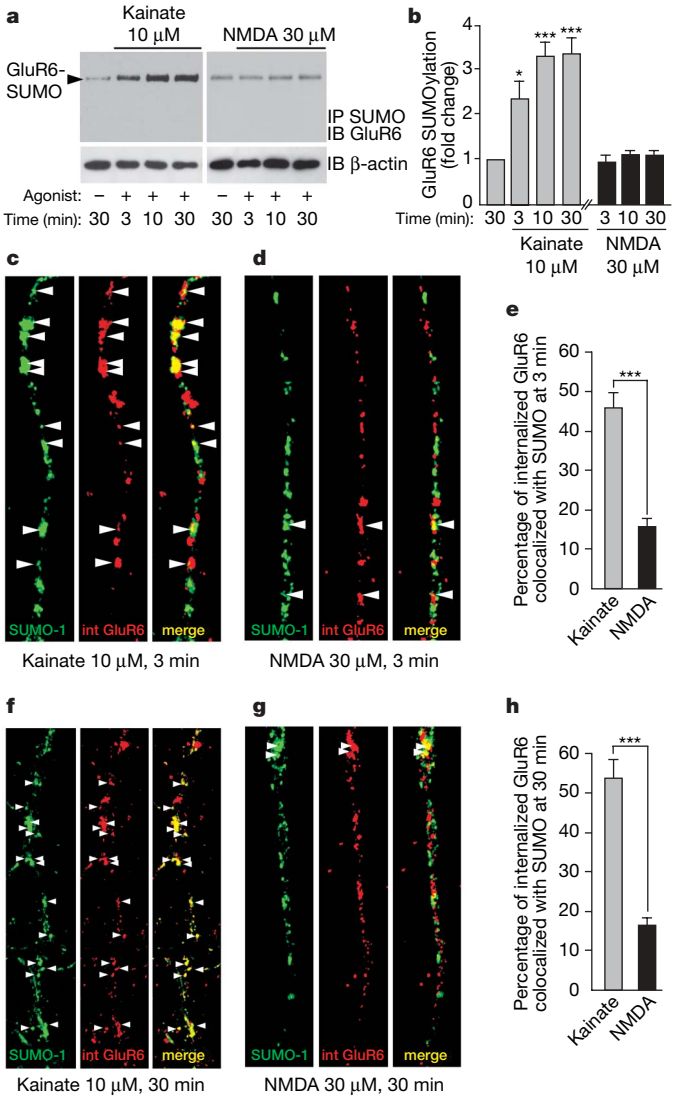


Figure 2 | GluR6a is SUMOylated after direct stimulation with kainate in cultured hippocampal neurons. a, Immunoprecipitation of SUMOylated protein and immunoblotting for GluR6 shows that 10 μM kainate application dramatically increases GluR6 SUMOylation in a time-dependent manner. In contrast, 30 μM NMDA does not change the SUMOylation state of GluR6. Lower panels show an immunoblot of total β-actin as a control of protein loading. **b**, Quantification of kainate and NMDA-induced GluR6 SUMOylation shown in **a**. The data are from four separate experiments and show mean ± s.e.m. $*P < 0.01$ and $***P < 0.001$ compared with unstimulated controls. **c–e**, Colocalization of SUMO-1 and endocytosed GluR6 3 min after 10 μM kainate (**c**) or 30 μM NMDA (**d**) stimulation. A significant proportion of internalized GluR6 (red) colocalizes as indicated by arrowheads (yellow) with SUMO-1 (green) after 3 min kainate stimulation, whereas few internalized ('int') receptors colocalize with SUMO-1 upon NMDA application, as quantified in **e**. Data show mean ± s.e.m. of three independent experiments. $***P < 0.0001$ compared with NMDA stimulation (*t*-test). **f–h**, Colocalization of SUMO-1 and endocytosed GluR6 30 min after 10 μM kainate (**f**) or 30 μM NMDA (**g**) stimulation. A large proportion of GluR6 internalized (red) in response to kainate stimulation colocalizes (yellow; arrowheads) with the SUMO labelling (green) whereas few internalized receptors colocalize with SUMO when neurons are stimulated with NMDA. **h**, Quantification of the percentage of internalized GluR6 that colocalizes with SUMO-1 in response to either kainate or NMDA receptor activation. Data show mean ± s.e.m. of three independent experiments. $***P < 0.0001$ compared with NMDA stimulation (*t*-test).

To identify a physiological role for GluR6 SUMOylation we investigated postsynaptic KARs at the mossy fibre synapse[13–15]. Under basal conditions stable, pharmacologically isolated, KAR-mediated, synaptically evoked excitatory postsynaptic currents (KAR-EPSCs) were recorded from CA3 neurons in hippocampal slices (Fig. 4a). Infusion of SUMO-1 into the CA3 neuron caused a rapid reduction in KAR-EPSC amplitude (97 ± 5% and 77 ± 8%; control versus SUMO-1; $P < 0.05$). The conjugation-inactive mutant SUMO-1-ΔGG had no effect (100 ± 6% and 77 ± 6%; SUMO-1-ΔGG versus SUMO-1; $P < 0.05$) (Fig. 4a, b, g). Infusion of recombinant SENP-1 increased the KAR-EPSC (98 ± 11% and 145 ± 13%; control versus SENP-1; $P < 0.05$). Infusion of the inactive SENP-1 C603S mutant had no effect (91 ± 11% and 145 ± 21%; SENP-1 C603S versus SENP-1; $P < 0.05$) (Fig. 4c, d, g). AMPA-receptor-mediated EPSCs recorded at the same synapse were unaffected by SUMO-1 or SENP-1 (89 ± 9% and 101 ± 13%; control versus SUMO-1; 101 ± 18% and 91 ± 13%; control versus SENP-1; Fig. 4e–g) or at CA1 Schaffer collateral synapses (Supplementary Fig. 11). KAR-EPSC decay kinetics were also unchanged by SUMO-1 ($\tau = 30.3 \pm 2.1$ ms during first 2 min of recording and 28.9 ± 1.7 ms during last 5 min, $n = 11$) or SENP-1 ($\tau = 26.8 \pm 2.1$ ms during first 2 min of recording and 25.6 ± 1.6 ms during last 5 min, $n = 11$).

We also infused SUMO-1 for 10 min before synaptic stimulation to determine whether SUMO-1 conjugation alone is sufficient to induce a decrease in KAR-EPSC amplitude. After the 10 min delay the reduction in KAR-EPSC amplitude was similar to those recorded immediately after membrane rupture (Fig. 4h), indicating that this process is activity-dependent because both SUMO-1 and agonist binding are required for the EPSC depression.

Our data indicate that the coincident action of agonist binding and SUMOylation to surface GluR6a leads to the rapid endocytosis of KARs. The observation that SENP-1 causes an increase in KAR-EPSC amplitude implies that when the removal of KARs is inhibited there is exocytosis and lateral diffusion of KARs into the synapse, as shown for AMPA receptors[16,17]. Also, NMDA stimulated GluR6 endocytosis is SUMO-1 independent, so there must be multiple KAR endo- and exocytosis pathways.

SUMO-1 regulation of KARs may be important in regulating the response of neurons to glutamate release under conditions such as ischaemia. Interestingly, it has been shown that SUMO-1 messenger RNA levels are dramatically raised in ischaemic cells[18]. Because GluR6 is chronically downregulated by excitotoxic stress via a c-fos dependent mechanism[19], one attractive possibility is that upregulation of SUMO levels leads to a downregulation in KARs, thereby reducing neuronal excitability. The mechanisms by which SUMOylation of GluR6 leads to internalization are unclear, but it seems likely that regulation of interacting proteins[20,21] may be involved.

Our findings, taken together with the diverse range of functions that SUMO serves in the nucleus and cytoplasm, suggest that this covalent protein modification is likely to regulate the functions of many different synaptic proteins. We anticipate that synaptic

SUMOylation represents a novel mechanism that will have far-reaching implications for understanding synaptic function.

METHODS SUMMARY

Yeast two-hybrid screen. Rat GluR6a C-terminal (841E–908A) domain was used as a bait to screen the rat whole-brain complementary DNA library.

Dissociated hippocampal cultures. Hippocampal cultures were prepared as previously reported[20] and plated at a density of 500,000 per 60 mm dish or 50,000 onto 22 mm glass coverslips coated with poly-L-lysine.

Biotinylation of cell surface protein. Live hippocampal neurons (21–25 days *in vitro*) were surface biotinylated on ice using the membrane impermeant Sulfo-NHS-SS-biotin.

Fluorescence imaging of GluR6. Live cultured hippocampal neurons (21–25 days *in vitro*) were transduced with Sinbdis to express appropriate protein constructs. Immunocytochemical imaging of surface and internalized receptors was performed on live and fixed cells using a Zeiss Meta confocal microscope.

Electrophysiology. Hippocampal slices were prepared from p14 Wistar rats. Whole-cell patch clamp recordings were made from visually identified CA3 pyramidal neurons and synaptic responses were recorded within 30 s of entering the whole-cell configuration.

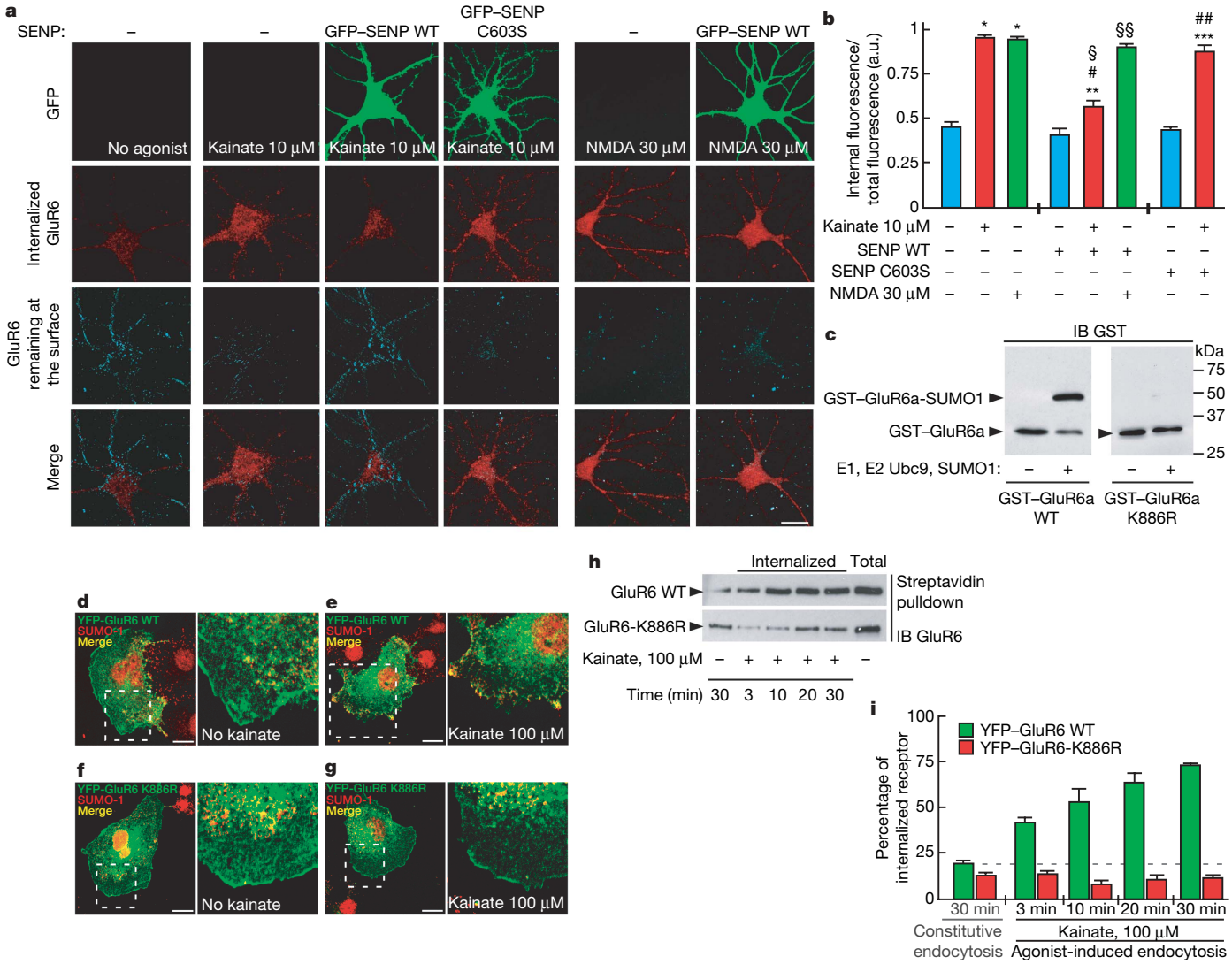


Figure 3 | DeSUMOylation blocks kainate-induced GluR6 endocytosis. **a**, Effect of kainate (10 μM) and NMDA (30 μM) application for 30 min at 37 °C on the endocytosis of GluR6 assessed by immunofluorescence assays[9] (see Methods). Internalized GluR6 is shown in red and the remaining labelled surface receptors are visualized in blue. The green labelling represents the fluorescence due to the GFP expression from hippocampal neurons transduced for 24 h with sindbis virus expressing either the active form of SENP-1 (GFP–SENP-1-WT) or its inactive point mutant form (GFP–SENP-1-C603S). SENP-1 overexpression does not prevent GluR6 endocytosis induced by NMDA stimulation, indicating a specific involvement of the SUMOylation pathway in ligand-induced KAR endocytosis. Scale bar, 20 μm. **b**, Quantification of GluR6 endocytosis expressed in arbitrary units (a.u.) corresponding to the ratio between the internalized and total GluR6 labelling as shown in **a**. The data are from three separate experiments and show mean ± s.e.m. $*P < 0.001$ compared with untreated uninfected neurons; $**P < 0.01$ compared with untreated GFP–SENP-1-WT; $***P < 0.001$ compared with untreated GFP–SENP-1-C603S; $\#P < 0.001$ compared with kainate-stimulated uninfected neurons; $\S P < 0.001$ compared with kainate-stimulated GFP–SENP-1-C603S; $\S\S P < 0.001$ compared with untreated GFP–SENP-1-WT; $\#\#P < 0.001$ compared with untreated uninfected neurons. **c**, Bacterial SUMOylation assay. The C terminus of GluR6a but not GluR6a K886R is SUMOylated in a recombinant bacterial system (see Methods for details). **d**–**g**, YFP-tagged GluR6a wild-type (**d**, **e**) and GluR6a-K886R point mutant subunits (**e**, **f**) were transiently transfected in COS-7 cells and stained with anti-SUMO-1 antibody (red). Cells were exposed to 100 μM kainate at 37 °C for 10 min (**e**, **g**) to trigger endocytosis. Untreated control cells are shown in **d** and **f**. Right panels in **d**–**g** are enlargements of the hatched boxes in the left panels of **d**–**g**. Scale bar, 10 μm. **h**, Representative immunoblots showing the time course of YFP–GluR6a wild-type and YFP–GluR6a-K886R mutant receptor endocytosis following kainate stimulation. Experiments were performed on transiently transfected COS-7 cells using biotinylation assays (see Methods for details). **i**, Time course of kainate-induced GluR6a wild-type and K886R point-mutant receptor endocytosis measured by biotinylation assay from **h**. Each time point represents the mean ± s.e.m. of four independent experiments.

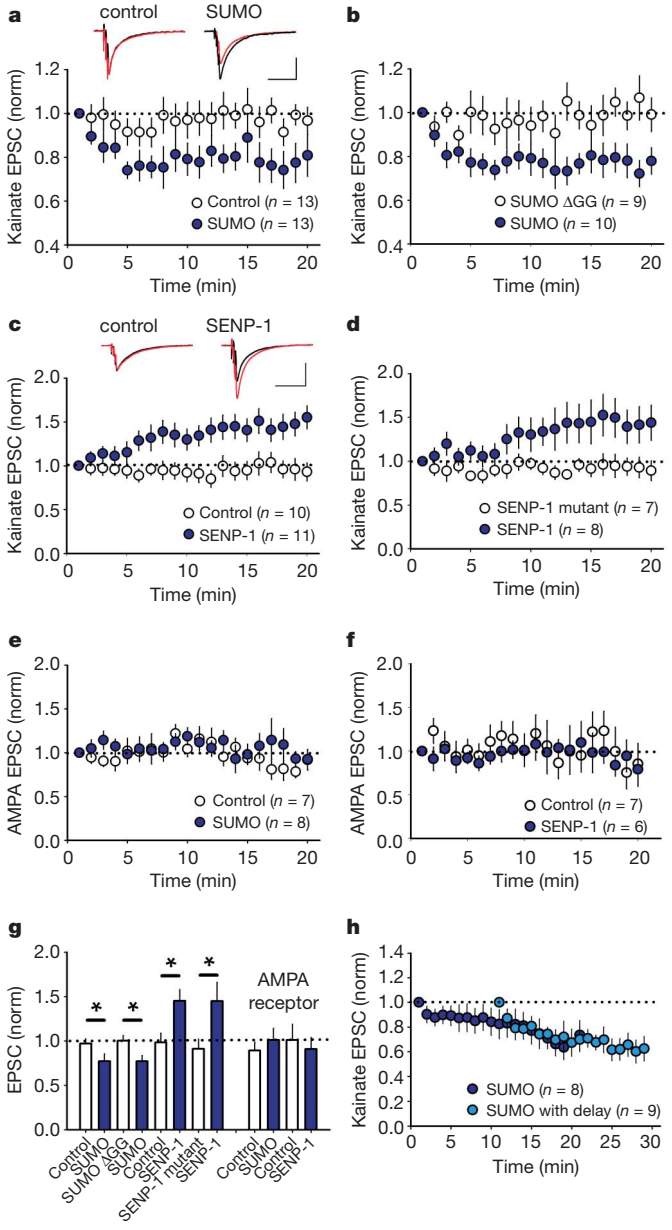


Figure 4 | Synaptic KARs are regulated by SUMOylation. **a**, Inclusion of 4.2 μM active SUMO-1 protein into the patch pipette caused a significant rundown in the KAR-mediated EPSC compared to interleaved control experiments. Responses for each cell are normalized to the first minute after membrane rupture. Example traces are taken as the average from the first minute (black) and 15–20 min (red). Scale bars are 50 pA (vertical) and 100 ms (horizontal). **b**, Same as **a** but the control contains 4.2 μM of the conjugation-inactive mutant SUMO-ΔGG. **c**, Inclusion of 100 nM SENP-1 caused a significant increase in the EPSC compared to interleaved control experiments. Example traces are taken as the average from the first minute (black) and 15–20 min (red). Scale bars are 50 pA and 100 ms. **d**, Same as **c** but the internal control contains 100 nM of the inactive mutant SENP-1-C603S. **e**, AMPA-receptor-mediated EPSCs were unaffected by the inclusion of 4.2 μM active SUMO. **f**, AMPA receptor-mediated EPSCs were unaffected by the inclusion of 100 nM SENP-1. **g**, Quantification of results shown in **a**–**e**. Statistical significance denoted by * ($P < 0.05$, unpaired *t*-test). **h**, A 10 min delay in commencing synaptic stimulation caused a similar rundown in kainate EPSC to interleaved control experiments, where in both cases the pipette contained 4.2 μM active SUMO (69 ± 10% versus 61 ± 9% 15–20 min after start of stimulation). Norm, normalized. All error bars are means ± s.e.m.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 December 2006; accepted 7 March 2007.
Published online 7 May 2007.

1. Seeler, J. S. & Dejean, A. Nuclear and unclear functions of SUMO. *Nature Rev. Mol. Cell Biol.* **4,** 690–699 (2003).
2. Wilson, V. G. & Rosas-Acosta, G. Wrestling with SUMO in a new arena. *Sci. STKE* **2005,** pe32 (2005).
3. Lerma, J. Roles and rules of kainate receptors in synaptic transmission. *Nature Rev. Neurosci.* **4,** 481–495 (2003).
4. Jaskolski, F., Coussen, F. & Mulle, C. Subcellular localization and trafficking of kainate receptors. *Trends Pharmacol. Sci.* **26,** 20–26 (2005).
5. Hilgarth, R. S. *et al.* Regulation and function of SUMO modification. *J. Biol. Chem.* **279,** 53899–53902 (2004).
6. Johnson, E. S. Protein modification by SUMO. *Annu. Rev. Biochem.* **73,** 355–382 (2004).
7. Suzuki, T. *et al.* A new 30-kDa ubiquitin-related SUMO-1 hydrolase from bovine brain. *J. Biol. Chem.* **274,** 31131–31134 (1999).
8. Hay, R. T. SUMO: a history of modification. *Mol. Cell* **18,** 1–12 (2005).
9. Martin, S. & Henley, J. M. Activity-dependent endocytic sorting of kainate receptors to recycling or degradation pathways. *EMBO J.* **23,** 4749–4759 (2004).
10. Gong, L., Millas, S., Maul, G. G. & Yeh, E. T. Differential regulation of sentrinized proteins by a novel sentrin-specific protease. *J. Biol. Chem.* **275,** 3355–3359 (2000).
11. Rajan, S., Plant, L. D., Rabin, M. L., Butler, M. H. & Goldstein, S. A. Sumoylation silences the plasma membrane leak K+ channel K2P1. *Cell* **121,** 37–47 (2005).
12. Uchimura, Y., Nakao, M. & Saitoh, H. Generation of SUMO-1 modified proteins in *E. coli*: towards understanding the biochemistry/structural biology of the SUMO-1 pathway. *FEBS Lett.* **564,** 85–90 (2004).
13. Vignes, M. & Collingridge, G. L. The synaptic activation of kainate receptors. *Nature* **388,** 179–182 (1997).
14. Castillo, P. E., Malenka, R. C. & Nicoll, R. A. Kainate receptors mediate a slow postsynaptic current in hippocampal CA3 neurons. *Nature* **388,** 182–188 (1997).
15. Mulle, C. *et al.* Altered synaptic physiology and reduced susceptibility to kainate-induced seizures in GluR6-deficient mice. *Nature* **392,** 601–605 (1998).
16. Luscher, C. *et al.* Role of AMPA receptor cycling in synaptic transmission and plasticity. *Neuron* **24,** 649–658 (1999).
17. Borgdorff, A. J. & Choquet, D. Regulation of AMPA receptor lateral movements. *Nature* **417,** 649–653 (2002).
18. Comerford, K. M. *et al.* Small ubiquitin-related modifier-1 modification mediates resolution of CREB-dependent responses to hypoxia. *Proc. Natl Acad. Sci. USA* **100,** 986–991 (2003).
19. Zhang, J. *et al.* c-fos regulates neuronal excitability and survival. *Nature Genet.* **30,** 416–420 (2002).
20. Hirbec, H. *et al.* Rapid and differential regulation of AMPA and kainate receptors at hippocampal mossy fibre synapses by PICK1 and GRIP. *Neuron* **37,** 625–638 (2003).
21. Coussen, F. *et al.* Co-assembly of two GluR6 kainate receptor splice variants within a functional protein complex. *Neuron* **47,** 555–566 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the Wellcome Trust (J.M.H.), the MRC (J.R.M. and J.M.H.) and the EU (J.M.H. from a GRIPPANT contract) for financial support. We thank G. Hodgkinson for validation electrophysiology experiments on recombinant GluR6 in HEK cells, and S. Correa for some preliminary immuno cytochemistry experiments. We also thank M. Fleck for anti-GluR6 antibody, S. Goldstein for the SENP-1 constructs, C. Mulle for the pcDNA3-myc-GluR6 and H. Saitoh for the bacterial SUMOylation system. We are grateful to M. Ashby, G. Banting, Z. Bashir, T. Bouschet, G. Collingridge, J. Hanley, J. Isaac, F. Jaskolski, A. Randall and D. Stephens for commenting on the manuscript.

Author Contributions S.M. and A. N. are co-first authors; J.R.M. and J.M.H. are co-last authors. S.M. performed surface expression, biochemistry and cell imaging assays in cell culture. A.N. made the original observation that GluR6a is a SUMOylation substrate, performed biochemistry on brain extracts and performed all molecular biological experiments. J.R.M. performed the electrophysiology experiments and co-wrote the manuscript. J.M.H. provided team leadership, project management and wrote the manuscript. All authors contributed to hypothesis development, experimental design and data interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.M.H. (j.m.henley@bris.ac.uk).

METHODS

Plasmids. A full list of the plasmids used in this study is provided in the Supplementary Notes under 'Plasmids'. Rat GluR6 C-terminal domain (accession number NM_019309) was subcloned into pBTM116. Truncation series were prepared by polymerase chain reaction (PCR). Human SUMO-1 (accession number NM_003352) was amplified from a brain cDNA library (Invitrogen). The His6–SUMO-1 fragment was produced by amplifying with a forward primer containing a *Bam*HI site before the start codon and a reverse primer inserting a stop codon followed by a *Xho*I site. Amplified fragments were digested and inserted into the pET30a(+) (Novagen) vector. pET30-His6–S-tag–SUMO-1–GST was prepared by inserting the full-length GST fragment amplified from pGEX4T-1 (Amersham Biosciences) after removing the stop codon from pET30-SUMO-1. The pMax(+)-SENP-1 (accession number BC045639) (both wild-type and C603S mutant) was a gift from S. Goldstein. GST–SENP-1-wild-type and GST–SENP-1-C603S were constructed by inserting SENP-1 between the *Bam*HI and *Sal*I sites in pGEX4T-1. The SENP-1 fragment, containing an entire isopeptidase catalytic domain, was excised as a *Xho*I and *Eco*RI fragment and inserted into pEGFP-C1 (Clontech) vector. GFP–SENP fusion constructs were then excised from pEGFP–SENP-1-wild-type/C603S between *Nhe*I and *Bsp*120I and inserted into the *Xba*I and *Bsp*120I site of pSinRep5(nsP2Ser) vector. Site-directed mutagenesis was performed using the QuickChange II or QuickChange II XL (Stratagene) systems according to the manufacturer's protocol. The integrity of all cDNA constructs was verified by DNA sequencing. To prepare the plasmid pcDNA3.2-YFP–myc6–GluR6 (wild-type and K886R) constructs, YFP–myc6–GluR6 (wild-type) was made by inserting the YFP fragment into the *Cla*I site at the 5′ side of the myc6 tag of pcDNA3-myc-GluR6 (gift from C. Mulle). After a new *Not*I site was created at the 3′ side of the stop codon by site-directed mutagenesis, 3′-UTR was removed as a *Not*I–*Not*I fragment. The K886R mutation was introduced by site-directed mutagenesis. The *Asp*718I-*Not*I insert fragment was subcloned into the same sites of the entry plasmid pENTR1A (Invitrogen). The expression plasmids were generated from site-specific recombination between these entry plasmids and pcDNA3.2/V5-DEST (Invitrogen) by using LR Clonase (Invitrogen) site-specific recombinase according to the manufacturer's protocol.

Yeast two-hybrid screen. L40 yeast strain bearing the C-terminal GluR6 bait plasmid pBTM116-GluR6(841–908) were transformed with the oligo-dT and random hexamer-primed rat whole-brain pGAD10 cDNA library (Clontech). Yeast colonies that grew in less than 4 days at 30 °C under selective conditions were streaked on SD-Trp-Leu-His and subjected to the β-galactosidase assay. Screens were performed as described previously[22]. Briefly, library plasmids containing rat-brain cDNAs were rescued from yeast and into *E. coli* strain HB 101, which contains the leuB mutation. Transformants were selected by ampicillin resistance and leucine autotrophy. Rescued plasmids were digested with *Eco*RI and classified according to their restriction patterns. Each representative plasmid was re-transformed into both L40 bearing the pBTM-GluR6 and L40 bearing the control bait plasmid pBTM-LAM. Yeast were plated on SD-Trp-Leu-His and subjected to the filter β-galactosidase assays to confirm the reproducibility and specificity of the interactions. The identities of the rat cDNA encoding Ubc9 and PIAS3 were determined by sequencing using pGAD10 specific primer followed by BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) analysis.

Virus preparation and transduction of hippocampal cultured neurons. Attenuated sindbis virus (SINrep(nsP2S726)) was prepared and used as previously reported[23]. cRNA was synthesized with an mMESSAGE mMACHINE SP6 kit after vector linearization (Ambion). cRNA from the construct in pSinRep5(nsP2S726) and the defective helper plasmid (pDH-BB(tRNA/TE12)) were mixed and electroporated into baby hamster kidney (BHK21) cells. 48 hours after electroporation, the culture medium containing the pseudovirions was harvested, ultracentrifuged and the pellet resuspended in a minimum volume of Neurobasal medium (Gibco). Aliquots of virus were stored at −80 °C. Cultured hippocampal neurons were transduced at a multiplicity of infection (MOI) of 3 and returned to the incubator for 24 to 48 h before use.

Expression and purification of recombinant proteins. pET-SUMO1(1–97; active form), pET-SUMO(1–95; ΔGG-conjugation-deficient form), or pET-His6–S-tag–SUMO1–GST were introduced to BL21(DE3) for recombinant protein expression. pGEX-SENP-1-wild-type and pGEX-SENP-1-C603S were introduced to BL21 for expression. Cells are cultured to the early-log phase and protein expression was induced by adding isopropyl-β-D-thiogalactoside (IPTG) at 0.2 mM final concentration. Cells were harvested and lysed by sonication in 1% Triton X-100, 5 mg ml^{-1} lysozyme and 20 U ml^{-1} Benzonase (Novagen) containing PBS(−). Cleared lysate was prepared by centrifuged at 35,000*g* for 20 minutes at 4 °C. His6-tagged proteins were purified with TALON Co^{2+} affinity resin (Clontech) and GST fusion proteins were purified by using glutathione Sepharose 4B (Amersham Biosciences), according to the manufacturer's protocols. After affinity chromatography purification, eluates were extensively dialysed against PBS(−) and the buffer exchanged to the intracellular solution for patch pipette using NAP-5 Sephadex G-25 columns. Eluates were adjusted to 100× concentration stocks and stored at −80 °C until use.

SUMOylation assay in *E. coli*. The bacterial SUMOylation assay was performed as described previously[12].

***In vitro* assay of recombinant SENP.** The recombinant SENP substrate His6–S-tag–SUMO1–GST was purified by Co^{2+} affinity chromatography. Assays were performed in 10 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 1 mM dithiothreitol (DTT). Reactions were incubated for 2 h at 30 °C. Proteins were subjected to SDS–polyacrylamide gel electrophoresis (PAGE) and electrotransfer onto PVDF-membrane. After blocking, intact and cleaved bands were detected by HRP-conjugated S-protein (Novagen) and chemiluminescence substrate.

Preparation of brain fractions. Brains from adult Wistar rats (220 g) were rapidly removed and put into ice-cold phosphate-buffered saline (PBS). Whole brains were homogenized in 20 mM HEPES (pH 7.4) containing 0.32 M sucrose, 150 mM NaCl, 1 mM $MgCl_2$, 0.5 mM $CaCl_2$ and 20 mM NEM. Subcellular fractionations were obtained by differential centrifugation as previously described[24].

Dissociated hippocampal cultures. Hippocampal cultures were prepared using a modified published protocol[9]. Briefly, hippocampi from E18 Wistar rats were dissected and the neurons dissociated by enzymatic digestion with trypsin for 15 min and mechanical dissociation. Cells were then plated at a density of 500,000 per 35 mm dish or 50,000 onto 22 mm glass coverslips coated with poly-L-lysine (Sigma). The culture medium was composed of Neurobasal medium (Gibco) supplemented with horse serum (10%), B27 (Gibco) and 2 mM glutamine. On the second day, the media was changed to Neurobasal medium supplemented with B27 only and the neurons were then fed weekly with this glutamine-free medium until use (21–25 days *in vitro*).

Biotinylation of cell surface protein. Live hippocampal neurons (21–25 days *in vitro*) were surface biotinylated on ice using the membrane impermeant Sulfo-NHS-SS-Biotin (Pierce, 0.15 mg ml^{-1} in PBS). After three washes in TBS (Tris-HCl 25 mM, pH 7.4, NaCl 137 mM, KCl 5 mM, $CaCl_2$ 2.3 mM, $MgCl_2$ 0.5 mM, Na_2HPO_4 0.143 g l^{-1}) to quench the remaining free unbound biotin, the cells were lysed in extraction buffer (Tris-HCl 50 mM pH 7.5, NaCl 150 mM, EDTA 10 mM, Triton X-100 1%, SDS 0.1%, mammalian protease inhibitor cocktail 1% and NEM 20 mM). After sonication for 10 s and centrifugation (15,000*g* for 20 min at 4 °C) supernatants containing equal amount of protein were incubated with streptavidin beads to immunoprecipitate the surface-biotinylated proteins. After many washes in extraction buffer, proteins were eluted from the streptavidin beads by boiling in reducing sample buffer and then resolved by SDS-PAGE and immunoblotted using rabbit polyclonal antibodies against GluR1 (1/2,000, Upstate Biotechnology) and/or GluR6/7 (1/2,000, Upstate Biotechnology). Bands were quantified using ImageJ software (NIH) and normalized to the total receptor band intensity.

Internalization assay in COS-7 cells. Transiently transfected COS-7 cells were biotinylated as described above. After extensive washes, cells were subsequently incubated in the absence or in the presence of 100 μM kainate for the times indicated in the legend. The remaining surface biotin was removed with GSH buffer (pH 9, 2 × 20 min), then lysed and incubated with streptavidin beads to isolate internalized biotinylated proteins. After washing, proteins were eluted from the streptavidin beads, separated on SDS-PAGE and immunoblotted with anti-GluR6/7 antibody (1/2,000) as previously described[9].

Immunoprecipitation. 200–400 μg of solubilized protein prepared as above was incubated with 2–4 μg of mouse monoclonal anti-SUMO-1 (D-11) antibody (from Santa Cruz) or 1/50 (v/v) rabbit anti-SUMO-1 (from Cell Signaling) overnight at 4 °C and then with 50 μl protein G-agarose beads (Sigma) for 1–3 h at 4 °C. Immunoprecipitates were washed four times with lysis buffer and proteins were eluted from the beads by boiling in reducing sample buffer and then resolved by SDS-PAGE.

Fluorescence imaging of GluR6 endocytosis. Live cultured hippocampal neurons (21–25 days *in vitro*) were transduced with sindbis virus containing the GFP-tagged catalytic domain of SENP-1 (active SENP-WT or its inactive form SENP-C603S) for 24 to 48 h. The neurons were then surface labelled for 20 min at room temperature with a chicken N-terminal directed anti-GluR6 antibody (1/4,000)[9,25,26]. Neurons were washed three times in PBS and then incubated for 30 min at 37 °C in Earle's-Tris-HEPES buffer containing 10 μM kainate (Tocris). For NMDA treatment, we used a modified protocol originally named "Chem-LTD"[27]. Briefly, neurons were stimulated at 37 °C with NMDA (30 μM) in Earle's-Tris-HEPES buffer for 3 min, and incubated in the same buffer in the absence of NMDA for the times indicated in the results. After a rapid 5 min fixation step in paraformaldehyde (4%), the remaining cell surface receptors were labelled with a Cy5-conjugated anti-chicken antibody (Jackson ImmunoResearch laboratories) for 45 min at room temperature. Neurons were then fixed for a further 15 min with paraformaldehyde (4%), permeabilized

for 2 min in PBS containing Triton X-100 (0.2%) and labelled with a Cy3-conjugated anti-chicken antibody (Jackson ImmunoResearch Laboratories) for one hour to visualize internalized receptors. Confocal images were acquired with a Zeiss LSM 510 confocal microscope and quantified using Volocity Software (Improvision). Values corresponding to the internalized receptors (red) were divided by the total fluorescence values (red + blue) to control for cell density and expressed as arbitrary units (a.u.).

COS-7 cells: YFP–myc6–GluR6a-WT and YFP–myc6–GluR6-K886R were transiently transfected in COS-7 cells and surface receptors were live labelled using an Alexa Fluor 647 coupled anti-GFP antibody for 10 min at room temperature (22 °C). After extensive washes in Earle's buffer at room temperature, cells were incubated at 37 °C for 20 min in Earle's buffer containing 100 μM kainate prior to fixation with paraformaldehyde (PAF) 4% in PBS for 5 min and labelling of remaining surface receptors with an Alexa Fluor 555 anti-GFP antibody. COS-7 cells were then washed in Earle's buffer and further fixed in PAF 4% in PBS for 15 min before mounting in Mowiol (Sigma) and confocal visualization.

Colocalization of internalized GluR6 with SUMO-1. Internalization experiments were carried out as above until the paraformaldehyde fixation, except that neurons were stimulated for 3 min with kainate (10 μM) and D-AP5 (50 μM) or NMDA (30 μM) for the indicated times at 37 °C. Remaining antibodies at the cell surface were removed by 2 × 2 min washes with low-pH Earle's buffer (0.5 M NaCl, pH 4 with acetic acid) and fixed for 20 min with paraformaldehyde (4%) as previously described[9]. After permeabilization, cells were incubated with anti-SUMO-1 (rabbit 1/100; Santa Cruz) and labelled with a Cy3-conjugated anti-chicken antibody to visualize internalized GluR6 and Cy2- or Cy5-conjugated secondary antibodies as indicated. Puncta corresponding to the internalized GluR6 that colocalized with the SUMO labelling were quantified using the LSM510 version 3.2 software (Zeiss). Analysis was made from at least eight cells for each condition.

Statistical analysis. Statistical analyses were performed using Prism 4 software (GraphPad). Data are expressed as mean ± s.e.m. Unpaired Student's *t*-tests and one-way analysis of variance (ANOVA) were performed with a Newman-Keuls post-test for multiple comparison data sets when required.

Electrophysiology. Hippocampal slices were prepared from p14 Wistar rats as previously described[28]. For recording, slices were immersed in a submerged recording chamber perfused with solution containing 126 mM NaCl, 10 mM glucose, 26 mM $NaHCO_3$, 2.5 mM KCl, 1.2 mM $NaHPO_4$, 2.5 mM $CaCl_2$ and 1.3 mM $MgSO_4$ at 35 °C and saturated with 95% O_2, 5% CO_2. Whole-cell patch clamp recordings were made from visually identified CA3 pyramidal neurons using borosilicate glass pipettes of resistance 2–6 MΩ filled with a solution containing 117 mM $CsMeSO_4$, 8 mM NaCl, 10 mM HEPES, 5 mM QX-314Cl, 4 mM MgATP, 0.3 mM NaGTP, 0.2 mM EGTA, 0.1 mM bestatin and 0.1 mM leupeptin at pH 7.4 and 280 mOsm. Synaptic responses were recorded within 30 s of entering the whole-cell configuration and elicited using a bipolar tungsten electrode placed in the granule cell layer stimulated every 20 s with three pulses applied at 167 Hz. Kainate responses were recorded in 50 μM D-APV, 100 μM picrotoxin and 25–40 μM GYKI53655. For AMPA receptor recording, the concentration of $CaCl_2$ and $MgSO_4$ were raised to 4 mM each, GYKI53655 was omitted and stimulation was given as one pulse every 20 s. DCG-IV (2 μM) was applied at the end of 23 experiments, in each case depressing the EPSC by >80%.

22. Nishimune, A. *et al.* NSF binding to GluR2 regulates synaptic transmission. *Neuron* **21,** 87–97 (1998).
23. Kim, J. *et al.* Sindbis vector SINrep(nsP2S726): a tool for rapid heterologous expression with attenuated cytotoxicity in neurons. *J. Neurosci. Methods* **133,** 81–90 (2004).
24. Hirbec, H., Martin, S. & Henley, J. M. Syntenin is involved in the developmental regulation of neuronal membrane architecture. *Mol. Cell. Neurosci.* **28,** 737–746 (2005).
25. Fleck, M. W., Cornell, E. & Mah, S. J. Amino-acid residues involved in glutamate receptor 6 kainate receptor gating and desensitization. *J. Neurosci.* **23,** 1219–1227 (2003).
26. Mah, S. J., Cornell, E., Mitchell, N. A. & Fleck, M. W. Glutamate receptor trafficking: endoplasmic reticulum quality control involves ligand binding and receptor function. *J. Neurosci.* **25,** 2215–2225 (2005).
27. Lee, H. K., Kameyama, K., Huganir, R. L. & Bear, M. F. NMDA induces long-term synaptic depression and dephosphorylation of the GluR1 subunit of AMPA receptors in hippocampus. *Neuron* **21,** 1151–1162 (1998).
28. Daw, M. I. *et al.* PDZ proteins interacting with C-terminal GluR2/3 are involved in a PKC- dependent regulation of AMPA receptors at hippocampal synapses. *Neuron* **28,** 873–886 (2000).

LETTERS

Herpesvirus latency confers symbiotic protection from bacterial infection

Erik S. Barton[1]†, Douglas W. White[1,5], Jason S. Cathelyn[2], Kelly A. Brett-McClellan[1], Michael Engle[3], Michael S. Diamond[1,2,3], Virginia L. Miller[2,4] & Herbert W. Virgin IV[1,2]

All humans become infected with multiple herpesviruses during childhood. After clearance of acute infection, herpesviruses enter a dormant state known as latency. Latency persists for the life of the host and is presumed to be parasitic, as it leaves the individual at risk for subsequent viral reactivation and disease[1]. Here we show that herpesvirus latency also confers a surprising benefit to the host. Mice latently infected with either murine gammaherpesvirus 68 or murine cytomegalovirus, which are genetically highly similar to the human pathogens Epstein–Barr virus and human cytomegalovirus[2], respectively, are resistant to infection with the bacterial pathogens *Listeria monocytogenes* and *Yersinia pestis*. Latency-induced protection is not antigen specific but involves prolonged production of the antiviral cytokine interferon-γ and systemic activation of macrophages. Latency thereby upregulates the basal activation state of innate immunity against subsequent infections. We speculate that herpesvirus latency may also sculpt the immune response to self and environmental antigens through establishment of a polarized cytokine environment. Thus, whereas the immune evasion capabilities and lifelong persistence of herpesviruses are commonly viewed as solely pathogenic, our data suggest that latency is a symbiotic relationship with immune benefits for the host.

Murine gammaherpesvirus 68 (γHV68) undergoes brief lytic replication in a number of cell types, and establishes lifelong latency in memory B cells, macrophages and dendritic cells *in vivo*[3]. While investigating the role of interferons in controlling γHV68 reactivation after intranasal infection, we noted that peritoneal macrophages from latently infected mice were uniformly activated, displaying cytoplasmic vacuolization, membrane ruffling, increased size and upregulation of surface major histocompatibility complex (MHC) class II molecules (Fig. 1a–d; Supplementary Fig. 1). Peritoneal macrophage activation was prominent for at least two months after mucosal inoculation of γHV68 and clearance of lytic infection (data not shown). As fewer than 1 in 500 peritoneal macrophages are latently infected[4], macrophage activation was not a direct consequence of infection. In contrast to macrophages explanted from mock-infected mice (Fig. 1e), macrophages from latently infected mice were bactericidal, killing *L. monocytogenes* rapidly after uptake (Fig. 1f, 6 h after infection).

Because activated macrophages are an important innate defence against bacterial pathogens, we considered the hypothesis that herpesvirus latency might be associated with host resistance to infection. To test this hypothesis, we challenged mice with *L. monocytogenes*, a Gram-positive intracellular pathogen controlled *in vivo* by activated macrophages[5]. Acute γHV68 infection did not alter susceptibility to this pathogen (Fig. 2a). In contrast, latently infected mice were highly resistant to *L. monocytogenes* (Fig. 2b). Significant protection from *L. monocytogenes* was observed up to 3 months after γHV68 infection (Fig. 2c), the latest time point assessed. Protection correlated with ~100-fold reduction in bacterial burden in the spleen and liver (Fig. 2d, e). A point mutant of γHV68 that is capable of lytic infection but severely defective in establishment of latency (ORF73.stop)[6] did not protect mice from *L. monocytogenes* infection and did not stimulate chronic macrophage activation (Fig. 2d, e; Supplementary Fig. 1), demonstrating the importance of latent infection in this

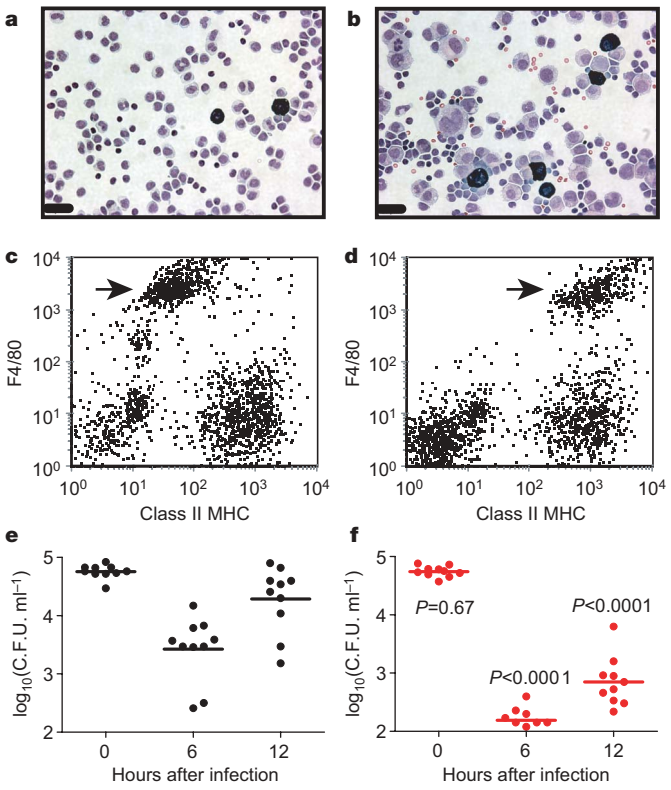


Figure 1 | Macrophages are activated during γHV68 latency. **a–f,** Peritoneal macrophages from mock (**a**, **c**, **e**) or latently infected (**b**, **d**, **f**) mice. **a**, **b**, Wright's Giemsa stain. Scale bar, 25 μm. **c**, **d**, Surface class II MHC levels on F4/80$^+$ peritoneal macrophages (arrow) in mock-infected mice (57 ± 14 (mean fluorescence intensity ± s.e.m.); $n = 23$) compared with latently infected mice (1,363 ± 144; $n = 18$). **e**, **f**, *Ex vivo* bactericidal activity of peritoneal macrophages at the indicated times. *P* values, calculated using the Mann–Whitney rank sum test, compare latently infected and mock-infected mice at the same time points. Data are representative of two to five independent experiments. c.f.u, colony-forming units. Horizontal bars indicate the arithmetic mean of log-transformed data.

[1]Departments of Pathology and Immunology, [2]Molecular Microbiology, [3]Medicine, and [4]Pediatrics, and [5]Division of Rheumatology, Washington University Medical School, 660 South Euclid Avenue, St. Louis, Missouri 63110, USA. †Present address: Purdue University, Department of Biological Sciences, 915 W. State Street, West Lafayette, Indiana 47907, USA.

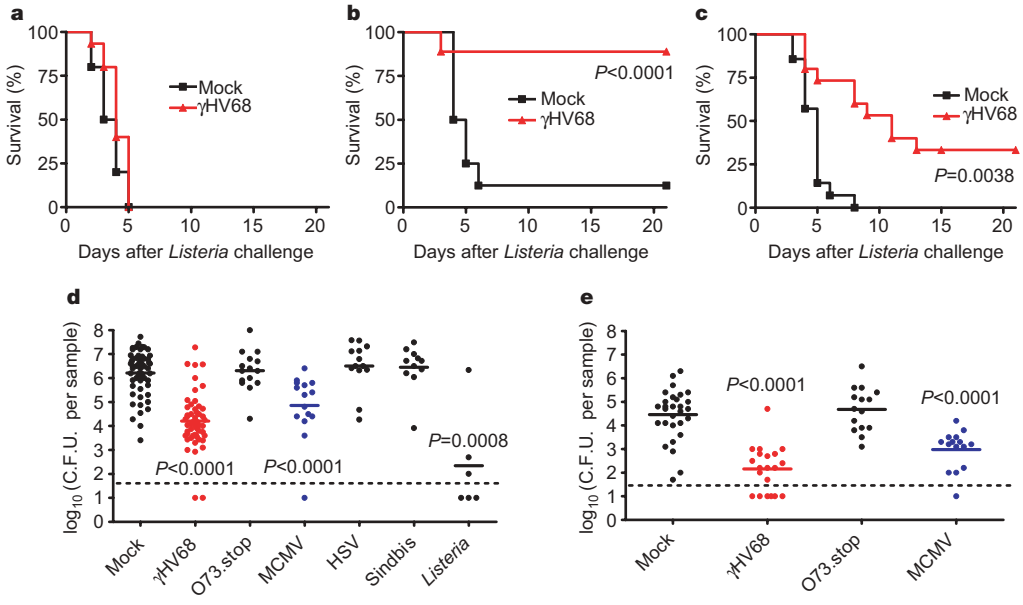


Figure 2 | Latent herpesvirus infection renders mice resistant to *L. monocytogenes*. **a**–**c**, Survival of mice infected with γHV68 1 (**a**), 4 (**b**) or 12 (**c**) weeks before challenge with *L. monocytogenes*. *P* values were calculated using logrank analysis (two to three experiments, $n = 8$–15 mice per condition). **d**, **e**, Bacterial titres in the spleen (**d**) and liver (**e**) 3 days after challenge in mice infected 28–42 days previously with the indicated virus or a sublethal dose of *L. monocytogenes*. *P* values were calculated using Mann–Whitney rank sum test compared to mock (three to five independent experiments). Horizontal bars indicate the arithmetic mean of log-transformed data. The dashed line represents the limit of detection. Where not indicated, $P > 0.3$.

protection. Latency-induced protection and macrophage activation were also observed in mice infected with murine cytomegalovirus (MCMV), a betaherpesvirus (Fig. 2d, e; Supplementary Fig. 1). Previous infection with human herpes simplex virus type-1 (HSV-1) or Sindbis virus (an RNA virus incapable of latent infection) triggered neither protection from bacterial infection nor chronic macrophage activation (Fig. 2d; Supplementary Fig. 1). Thus, latent infection with two divergent herpesviruses results in protection of the host from *L. monocytogenes*, indicating that this phenotype is a general and previously unappreciated consequence of herpesvirus latency.

To explore the possibility that herpesvirus latency enhanced antibacterial resistance more generally, we challenged γHV68 latently infected mice with the human and rodent pathogen *Y. pestis*, a Gram-negative extracellular bacterium and the causative agent of plague[7]. Using a pneumonic plague model, mock-infected mice displayed rapid *Y. pestis* replication in the lung, with spread to the spleen at later time points (Fig. 3a, b). In contrast, latently infected mice were resistant, exhibiting decreased *Y. pestis* replication in the lung (Fig. 3a) and decreased systemic spread (Fig. 3b). Latently infected mice also exhibited resistance in a bubonic plague model, displaying 100- to 1,000-fold decreases in systemic *Y. pestis* replication (Fig. 3c, d). The protection afforded by latency was not universal, as latently infected mice succumbed to West Nile virus (WNV) infection with kinetics indistinguishable from mock-infected mice (Fig. 3e).

To define the mechanism whereby herpesvirus latency protects against two distinct bacterial pathogens, we measured cytokine levels in the serum of latently infected mice. We found that latent infection triggered elevated levels of interferon-γ (IFN-γ) and tumour-necrosis factor-α (TNF-α), soluble mediators of macrophage activation and resistance to bacterial infection (Fig. 4a). IFN-γ was required for latency-induced protection from bacterial challenge (Fig. 4b), and consistent with a critical role for latent infection in this process, we did not observe IFN-γ upregulation in mice infected with the latency-defective ORF73.stop mutant (data not shown). Infection of $H\text{–}2K^{b-/-}$ $H\text{–}2D^{b-/-}$ $B_2M^{-/-}$ mice indicated that classical class I MHC-restricted $CD8^+$ T-cells, an important source of IFN-γ during γHV68 infection[8], were completely dispensable for protection (Fig. 4c). Notably, both $Ifng^{-/-}$ and $H\text{–}2K^{b-/-}$ $H\text{–}2D^{b-/-}$ $B_2M^{-/-}$ mice clear acute infection with γHV68 and

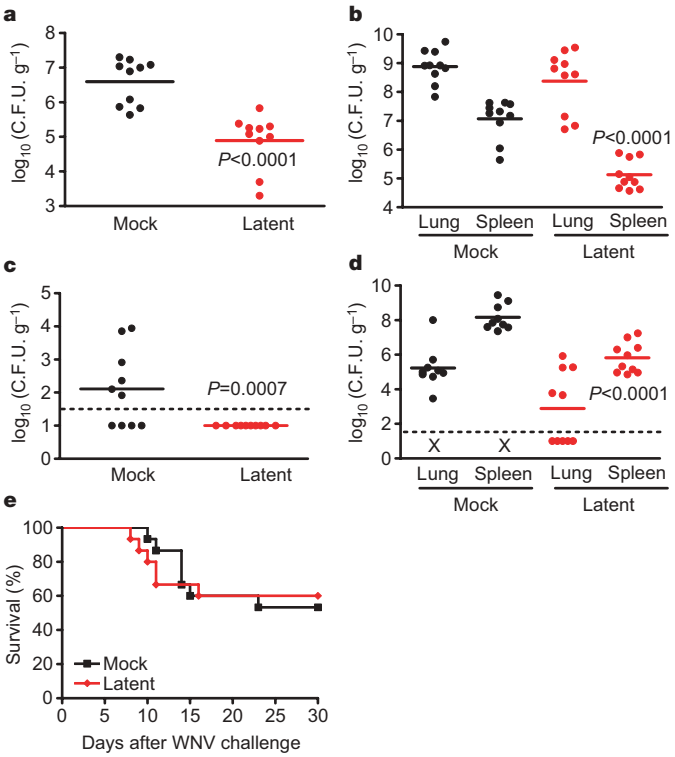


Figure 3 | Latently infected mice are resistant to *Y. pestis* but not WNV. **a**–**d**, Bacterial titres in mock (black symbols) or latently infected (γHV68, red symbols) mice challenged intranasally (**a**, 24 hours past infection (h.p.i.), lung, **b**, 48 h.p.i.) or subcutaneously (**c**, 36 h.p.i., spleen, **d**, 60 h.p.i.) with *Y. pestis*. *P* values, calculated using the Mann–Whitney rank sum test, compare latently infected and mock-infected mice (two independent experiments). X indicates death before harvest. Horizontal bars indicate the arithmetic mean of log-transformed data. The dashed line represents the limit of detection. Where not indicated, $P > 0.08$. **e**, Survival of mock or latently infected mice challenged by footpad inoculation with WNV. The *P* value was calculated using logrank analysis with ten mice per group.

establish latent infection, which is associated with persistent replication below the level of detection of standard plaque assays[8–10]. This indicates that latency-induced protection does not require classical $CD8^+$ T-cell-derived IFN-γ, and may be mediated by other cell types that have been demonstrated to produce IFN-γ during gammaherpesvirus latency, including activated $CD4^+$ T cells[3,11–13].

Infection-induced 'cross protection' between unrelated bacterial pathogens was the foundational observation defining cell-mediated immunity[14]. Classical cross protection occurs when effector lymphocytes responding to the initial infection secrete IFN-γ, thereby activating bystander macrophages and generating a heightened state of innate immunity to secondary infection. This classical form of cross protection wanes rapidly once the primary pathogen is cleared and responding lymphocytes differentiate from effector to memory phenotype[15], and has not previously been observed during latent virus infection.

In contrast, two mechanisms of cross protection that are mediated by memory lymphocytes have been recently described. In the first, termed heterologous immunity, cross-reactive antigenic epitopes between the primary and secondary pathogens result in antigen-specific memory lymphocyte activation after secondary infection[16]. In the second mechanism, $CD8^+$ memory T cells generated after primary infection can become activated in an antigen-independent manner by cytokines (including interleukin (IL)-12 and IL-18) expressed during the early stages of subsequent infections. In both cases, memory T-cell activation leads to secretion of IFN-γ and an enhanced response to the secondary infection[17]. In contrast to classical cross protection, memory-mediated cross protection does not involve heightened innate immunity before secondary challenge.

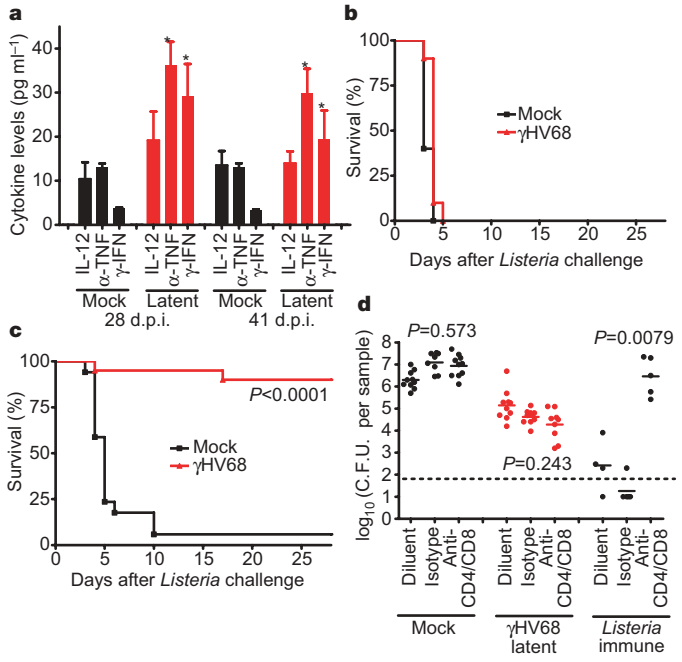


Figure 4 | Mechanism of enhanced innate immunity in latently infected mice. **a**, Cytokine levels (mean ± s.e.m.) following infection with γHV68. Asterisk, $P \leq 0.0002$, relative to mock, Mann–Whitney rank sum. Where not indicated, $P > 0.4$. **b**, **c**, Survival of mock or latently infected *Ifng*$^{-/-}$ (**b**) or *H–2K*$^{b-/-}$ *H–2D*$^{b-/-}$ *B*$_2$*M*$^{-/-}$ (**c**) mice, challenged with *L. monocytogenes*. The *P* value is relative to mock, logrank analysis (two to three experiments, $n = 10$–20 mice per condition). **d**, Bacterial titres in spleen 3 days after challenge with *L. monocytogenes* in mock, latently infected or *L. monocytogenes*-immune mice. $CD4^+$ and $CD8^+$ T cells were depleted 1 day before challenge. Horizontal bars indicate the arithmetic mean of log-transformed data. *P*-values, calculated using the Mann–Whitney rank sum test, compare depleted to isotype controls (two experiments).

Rather, the second pathogen activates an innate cytokine response that triggers bystander memory T-cell activation.

The chronic IFN-γ secretion and macrophage activation we observe in latently infected mice suggest that latency induces a prolonged state of classical cross protection. However, these observations do not exclude a role for the T-cell-dependent mechanisms described above. To determine whether latency-induced cross protection requires activation of memory T cells at the time of bacterial challenge, we depleted both $CD4^+$ and $CD8^+$ lymphocytes one day before *L. monocytogenes* challenge (Fig. 4d). T-cell depletion was effective, as indicated by elimination of *L. monocytogenes*-specific memory in depleted *L. monocytogenes*-vaccinated mice (Fig. 4d). In contrast, the resistance of latently infected mice to *L. monocytogenes* remained when $CD4^+$ and $CD8^+$ T cells were depleted. We conclude that viral latency generates a state of broadly effective innate cross protection that does not require the effector functions of lymphocytes at the time of secondary challenge.

Latency-induced cross protection differs from classical cross protection in at least two aspects. First, IFN-γ-dependent cross protection can occur between latent viral infection and acute bacterial infections. Second, this protection is quite prolonged, lasting months after clearance of acute infection. The durability of cross protection that we observe during gammaherpesvirus latency leads us to propose that latent infection represents a true symbiotic relationship, protecting the host from subsequent infection by means of heightened innate immune activation. Although our data suggest that IFN-γ-activated macrophages mediate latency-induced cross protection during both beta- and gammaherpesvirus infection, other cell types may contribute to the protective effects observed.

Are our findings with murine herpesviruses relevant for other latently infected hosts? We are unaware of studies that directly address the role of herpesvirus infection in human susceptibility to secondary bacterial infection. However, the observation of latency-induced cross protection in two subfamilies of the Herpesviridae suggests that this is an evolutionarily conserved aspect of the virus–host relationship. Furthermore, accumulating evidence indicates that latency with all three herpesvirus subfamilies in humans involves chronic, low-level immune activation accompanied by IFN-γ and TNF-α secretion in response to frequent but subclinical viral reactivation[18–22]. Although we do not observe latency-induced cross protection in mice infected with the alpha-herpesvirus HSV-1, this may reflect the tight anatomical restriction of HSV-1 latency in neurons, in contrast to latent infection of circulating haematopoietic cells by beta- and gammaherpesviruses. A plausible hypothesis for the mechanism of latency-induced cross protection involves the chronic, low-level presentation of viral antigens as a result of viral reactivation, resulting in prolonged T-cell activation and secretion of IFN-γ. This effect may be particularly important at mucosal and epithelial sites, where herpesvirus reactivation occurs and most pathogenic challenges initiate. Our data clearly indicate that IFN-γ expression during latency can have beneficial effects beyond control of viral reactivation, and suggest the need for further epidemiological studies of the immune consequences of herpesvirus infection in humans. These studies should also take into account the possible deleterious consequences of chronic inflammation elicited during latency, such as the development of autoimmune disease or cancer[23].

Beneficial immune modulation during herpesvirus latency may extend beyond protection from pathogens. Substantial epidemiological data support the 'hygiene' hypothesis, in which prior infection protects against development of allergy[24]. Few studies have addressed the role of herpesvirus latency in skewing immune responses, despite the increasing age of seroconversion to herpesvirus infection in populations with increased atopy and autoimmunity[25]. One study reports a protective effect of Epstein–Barr virus infection in the reduction of IgE sensitization to environmental allergens, an effect that was more pronounced on co-infection with human cytomegalovirus and Epstein–Barr virus[26]. Our findings suggest a mechanistic

explanation for this observation, whereby prolonged secretion of T-helper-1-type cytokines during latency may inhibit the development of T-helper-2-driven immune pathology. Accordingly, potential benefits of herpesvirus latency should be considered when evaluating vaccines against herpesvirus infection, as decreased infection rates may be associated with unintended negative consequences for vaccinated individuals.

Our current understanding of cellular and molecular immunity is founded largely on studies in specific-pathogen-free mice that have no known herpesvirus infections. We now demonstrate that herpesvirus infection triggers systemic, profound immune modulation, with the potential to alter significantly the kinetics and nature of host response to foreign antigens. As there are few herpesvirus-free humans, we speculate that an accurate understanding of 'normal' immunity may require a re-evaluation of immune function in the presence of these symbionts, which have been exerting selective pressures on the mammalian immune system since before the vertebrate radiation[27].

METHODS SUMMARY

Mice, viruses and infections. C57BL/6J mice were infected with virus between 8 and 20 weeks of age. *Ifng*$^{-/-}$ or *H–2K*$^{b-/-}$ *H–2D*$^{b-/-}$ *B*$_2$*M*$^{-/-}$ mice are described elsewhere[8,10]. Mice received wild-type or ORF73.stop γHV68 (WUMS strain, intranasal), HSV-1 (strain 17, intraperitoneal), MCMV (strain Smith, intraperitoneal), or Sindbis virus (AR339, intraperitoneal). Mice were considered latently infected at 28 days after infection. *Listeria monocytogenes* was administered intraperitoneally. *Yersinia pestis* was administered intranasally[7] or subcutaneously. WNV was administered by footpad inoculation[28]. Lymphocytes were depleted by administration of anti-CD4 (YTS-191.1), anti-CD8 (H35), or isotype control SFR-DR5 rat monoclonal antibodies intraperitoneally[29].

Flow cytometric analysis. Peritoneal cells were obtained by lavage, blocked in 10% normal mouse serum, and stained for cell-surface MHC class II (I–Ab/I–Eb; Phyroerythrin-conjugated, Pharmingen) and F4/80 (allophycocyanin-conjugated, Caltag) and analysed on a Beckton-Dickinson FACScalibur.

***L. monocytogenes* growth in macrophages.** Adherent peritoneal macrophages were infected with *L. monocytogenes*; unbound bacteria were removed. At 30 min after-adsorption, gentamicin (5 μg ml^{-1}) was added to kill extracellular bacteria, permitting the direct assessment of macrophage capacity to kill phagocytosed *L. monocytogenes*. Viable intracellular bacteria were quantified by hypotonic lysis of peritoneal cells and plating onto BHI agar plates[30].

Quantification of bacterial growth *in vivo*. Spleens, livers and lungs were removed from *L. monocytogenes*- or *Y. pestis*-infected mice at the indicated times after bacterial challenge. For quantification of *L. monocytogenes*, spleens were homogenized in 10 ml PBS/0.05% Triton X-100 and plated on BHI-agar plates. Organs from *Y. pestis*-infected mice were homogenized in sterile PBS and plated onto BHI agar plates.

Quantification of cytokine expression. Cytokines were detected in whole serum using the mouse inflammation cytometric bead array (BD Pharmingen) according to the manufacturer's instructions.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 February; accepted 16 March 2007.

1. Rickinson, A. B. & Kieff, E. *Fields Virology* (eds Knipe, D. M. & Howley, P.) 2655–2670 (Lippincott-Raven, Philadelphia, 2007).
2. Virgin, H. W. *et al.* Complete sequence and genomic analysis of murine gammaherpesvirus 68. *J. Virol.* **71,** 5894–5904 (1997).
3. McClellan, K. B., Gangappa, S., Speck, S. H. & Virgin, H. W. Antibody-independent control of gamma-herpesvirus latency via B cell induction of anti-viral T cell responses. *PLoS Pathog.* **2,** e58 (2006).
4. Weck, K. E., Kim, S. S., Virgin, H. W. & Speck, S. H. Macrophages are the major reservoir of latent murine gammaherpesvirus 68 in peritoneal cells. *J. Virol.* **73,** 3273–3283 (1999).
5. Edelson, B. T. & Unanue, E. R. Immunity to *Listeria* infection. *Curr. Opin. Immunol.* **12,** 425–431 (2000).
6. Moorman, N. J., Willer, D. O. & Speck, S. H. The gammaherpesvirus 68 latency-associated nuclear antigen homolog is critical for the establishment of splenic latency. *J. Virol.* **77,** 10295–10303 (2003).
7. Cathelyn, J. S., Crosby, S. D., Lathem, W. W., Goldman, W. E. & Miller, V. L. RovA, a global regulator of *Yersinia pestis*, specifically required for bubonic plague. *Proc. Natl Acad. Sci. USA* **103,** 13514–13519 (2006).
8. Braaten, D. C. *et al.* Effective control of chronic gamma-herpesvirus infection by unconventional MHC Class Ia-independent CD8 T cells. *PLoS Pathog.* **2,** e37 (2006).
9. Sarawar, S. R. *et al.* Gamma interferon is not essential for recovery from acute infection with murine gammaherpesvirus 68. *J. Virol.* **71,** 3916–3921 (1997).
10. Tibbetts, S. A., Van Dyk, L., Speck, S. H. & Virgin, H. W. Immune control of the number and reactivation phenotype of cells latently infected with a gamma-herpesvirus. *J. Virol.* **76,** 7125–7132 (2002).
11. Christensen, J. P. & Doherty, P. C. Quantitative analysis of the acute and long-term CD4$^+$ T-cell response to a persistent gammaherpesvirus. *J. Virol.* **73,** 4279–4283 (1999).
12. Flano, E., Woodland, D. L., Blackman, M. A. & Doherty, P. C. Analysis of virus-specific CD4$^+$ T cells during long-term gammaherpesvirus infection. *J. Virol.* **75,** 7744–7748 (2001).
13. Polic, B. *et al.* Hierarchical and redundant lymphocyte subset control precludes cytomegalovirus replication during latent infection. *J. Exp. Med.* **188,** 1047–1054 (1998).
14. Mackaness, G. B. The immunological basis of acquired cellular resistance. *J. Exp. Med.* **120,** 105–120 (1964).
15. Mackaness, G. B. The influence of immunologically committed lymphoid cells on macrophage activity *in vivo*. *J. Exp. Med.* **129,** 973–992 (1969).
16. Selin, L. K. *et al.* Memory of mice and men: CD8$^+$ T-cell cross-reactivity and heterologous immunity. *Immunol. Rev.* **211,** 164–181 (2006).
17. Berg, R. E., Crossley, E., Murray, S. & Forman, J. Memory CD8$^+$ T cells provide innate immune protection against *Listeria monocytogenes* in the absence of cognate antigen. *J. Exp. Med.* **198,** 1583–1593 (2003).
18. Tan, L. C. *et al.* A re-evaluation of the frequency of CD8$^+$ T cells specific for EBV in healthy virus carriers. *J. Immunol.* **162,** 1827–1835 (1999).
19. Dunn, H. S. *et al.* Dynamics of CD4 and CD8 T cell responses to cytomegalovirus in healthy human donors. *J. Infect. Dis.* **186,** 15–22 (2002).
20. Amyes, E. *et al.* Characterization of the CD4$^+$ T cell response to Epstein-Barr virus during primary and persistent infection. *J. Exp. Med.* **198,** 903–911 (2003).
21. Hislop, A. D. *et al.* Tonsillar homing of Epstein-Barr virus-specific CD8+ T cells and the virus-host balance. *J. Clin. Invest.* **115,** 2546–2555 (2005).
22. Theil, D. *et al.* Latent herpesvirus infection in human trigeminal ganglia causes chronic immune response. *Am. J. Pathol.* **163,** 2179–2184 (2003).
23. Balkwill, F. & Mantovani, A. Inflammation and cancer: back to Virchow? *Lancet* **357,** 539–545 (2001).
24. Schaub, B., Lauener, R. & von Mutius, E. The many faces of the hygiene hypothesis. *J. Allergy Clin. Immunol.* **117,** 969–977 (2006).
25. Takeuchi, K. *et al.* Prevalence of Epstein-Barr virus in Japan: trends and future prediction. *Pathol. Int.* **56,** 112–116 (2006).
26. Nilsson, C. *et al.* Does early EBV infection protect against IgE sensitization? *J. Allergy Clin. Immunol.* **116,** 438–444 (2005).
27. Roizman, B. & Pellett, P. E. *Fields Virology* (eds Knipe, D. M. & Howley, P. M.) 2475–2499 (Lippincott-Raven, Philadelphia, 2007).
28. Shrestha, B. *et al.* Gamma interferon plays a crucial early antiviral role in protection against West Nile virus infection. *J. Virol.* **80,** 5338–5348 (2006).
29. McClellan, J. S., Tibbetts, S. A., Gangappa, S., Brett, K. A. & Virgin, H. W. Critical role of CD4 T cells in an antibody-independent mechanism of vaccination against gamma-herpesvirus latency. *J. Virol.* **78,** 6836–6845 (2004).
30. Edelson, B. T. & Unanue, E. R. Intracellular antibody neutralizes Listeria growth. *Immunity* **14,** 503–512 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by grants from the National Institutes of Health (E.S.B., V.L.M. & H.W.V.), an Abbott Scholar Award (D.W.W.) and a Cancer Research Institute postdoctoral fellowship (E.S.B.).

Author Contributions The original hypothesis of the article was formulated by E.S.B. and H.W.V. E.S.B. and D.W.W performed all experiments except those characterizing *Y. pestis* infection (J.S.C. and V.L.M.) and WNV infection (M.E. and M.S.D.). K.A.B.-M. made the initial observation of chronic IFNγ secretion during latency. The manuscript was written by E.S.B., D.W.W., and H.W.V. and all authors commented on data and conclusions prior to submission.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.W.V. (virgin@wustl.edu).

METHODS

Mice and infections. C57BL/6J mice (Jackson and internal breeding colonies) were infected with virus between 8 and 20 weeks of age. *Ifng*$^{-/-}$ or *H–2K*$^{b-/-}$ *H–2D*$^{b-/-}$ $B_2M^{-/-}$ mice are described elsewhere[8,10] and were from internal breeding colonies. γHV68 infections (10^4 plaque-forming units (p.f.u.)) were performed intranasally in a volume of 40 μl DMEM/10% FBS (DMEM/10). Lytic infection of ORF73.stop virus after intranasal inoculation was confirmed by enzyme-linked immunosorbent assay (ELISA) against total γHV68 antigen, which demonstrated equivalent serum levels of γHV68-specific antibody after wild-type γHV68 or ORF73.stop infection (data not shown). HSV1 (10^4 p.f.u.), MCMV (10^5 p.f.u.), and Sindbis virus (10^4 p.f.u.) were administered intraperitoneally in 0.5 ml DMEM/10. *Listeria monocytogenes* (nonlethal dose, 10^3 c.f.u; lethal dose 10^6 c.f.u) was administered intraperitoneally in 0.5 ml saline. MCMV infected mice displayed no serological evidence of γHV68 infection by ELISA (data not shown). *Yersinia pestis* used for subcutaneous infections were cultured at 26 °C in BHI for 16–18 h, and for intranasal infections bacteria were cultured at 37 °C in BHI supplemented with 2.5 mM $CaCl_2$ for 16–18 h. For analysis of *Y. pestis* colonization and dissemination mice were infected intranasally with 1×10^4 CFU essentially as described[7] or subcutaneously with 1×10^2 c.f.u. in 100 μl at the base of the tail. WNV (100 p.f.u.) was administered by footpad inoculation as described[28]. In some experiments, lymphocytes were depleted by administration of a single 1-mg dose of anti-CD4 (YTS-191.1), anti-CD8 (H35), or isotype control SFR-DR5 (American Type Culture Collection) rat monoclonal antibody intraperitoneally[29].

Viral and bacterial stocks. γHV68 (strain WUMS) stocks were prepared and titrated in murine 3T12 fibroblasts. The latency-deficient γHV68 mutant ORF73.stop was obtained from the laboratory of S. H. Speck[6]. MCMV strain Smith was obtained from the laboratory of M. Colonna. HSV-1 strain 17 and Sindbis virus strain AR339 are laboratory stocks. *Listeria monocytogenes* strain EGD stocks were amplified from infected splenic lysate and were obtained from the laboratory of E. R. Unanue. The virulent wild-type *Y. pestis* strain CO92[7] was obtained from the US Army, Ft. Detrick, Maryland. WNV (strain 3000.0259) was isolated in New York in 2000 and passaged once in C6/36 *Aedes albopictus* cells[28].

Quantification of *L. monocytogenes* growth in cultured peritoneal macrophages. This assay has been described in detail previously[30]. Briefly, peritoneal cells were removed by lavage with 10 ml ice-cold DMEM/10 and allowed to adhere to polystyrene 24-well plates overnight. Non-adherent cells were removed by washing with agitation in warm DMEM/10. Monolayers were infected with 1×10^5 c.f.u. of fresh-standing overnight culture of *L. monocytogenes* by centrifugal adsorption. Unbound bacteria were removed by rinsing with warm PBS and then cultured in the presence of gentamicin (5 μg ml^{-1}, added at 30 min after bacterial adsorption) to kill extracellular bacteria that have not been phagocytosed, thereby permitting the direct assessment of macrophage capacity to kill phagocytosed *L. monocytogenes* before bacterial escape from the endosome. At various times after adsorption, viable intracellular bacteria were quantified by hypotonic lysis of macrophage monolayers in water and inoculation of serial dilutions onto BHI agar plates.

Vol 447|17 May 2007|doi:10.1038/nature05765

LETTERS

Helicobacter pylori CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity

Iraj Saadat[1], Hideaki Higashi[1], Chikashi Obuse[2], Mayumi Umeda[1], Naoko Murata-Kamiya[1], Yasuhiro Saito[1], Huaisheng Lu[1], Naomi Ohnishi[1], Takeshi Azuma[3], Atsushi Suzuki[4], Shigeo Ohno[4] & Masanori Hatakeyama[1]

***Helicobacter pylori cagA*-positive strains are associated with gastritis, ulcerations and gastric adenocarcinoma[1]. CagA is delivered into gastric epithelial cells[2] and, on tyrosine phosphorylation, specifically binds and activates the SHP2 oncoprotein[3–7], thereby inducing the formation of an elongated cell shape known as the 'hummingbird' phenotype[2,3]. In polarized epithelial cells, CagA also disrupts the tight junction and causes loss of apical–basolateral polarity[8,9]. We show here that *H. pylori* CagA specifically interacts with PAR1/MARK kinase, which has an essential role in epithelial cell polarity[10,11]. Association of CagA inhibits PAR1 kinase activity and prevents atypical protein kinase C (aPKC)-mediated PAR1 phosphorylation, which dissociates PAR1 from the membrane[12,13], collectively causing junctional and polarity defects. Because of the multimeric nature of PAR1 (ref. 14), PAR1 also promotes CagA multimerization, which stabilizes the CagA–SHP2 interaction[15]. Furthermore, induction of the hummingbird phenotype by CagA-activated SHP2 requires simultaneous inhibition of PAR1 kinase activity by CagA. Thus, the CagA–PAR1 interaction not only elicits the junctional and polarity defects but also promotes the morphogenetic activity of CagA. Our findings revealed that PAR1 is a key target of *H. pylori* CagA in the disorganization of gastric epithelial architecture underlying mucosal damage, inflammation and carcinogenesis.**

The tight junction controls paracellular permeability across the epithelial cell monolayer. It also has a fundamental role in the establishment and maintenance of epithelial cell polarity by delimiting the apical and the basolateral membrane domains[16]. Expression of wild-type CagA (ABCCC-type Western CagA)[3,4] in polarized Madin–Darby canine kidney (MDCK) epithelial cells elicited mislocalization of the tight junction marker ZO-1 as previously reported[8,9] (Supplementary Fig. 1a). This observation is consistent with reduced ZO-1 staining at the tight junctions in gastric mucosa of patients infected with *cagA*-positive *H. pylori* (Supplementary Fig. 1b) and is also consistent with the results of a study showing that the tight junction structure was irregular and fragmented in *H. pylori*-associated gastritis mucosa[17]. A phosphorylation-resistant CagA (PR CagA)[3] (Supplementary Fig. 2) retained the ability to mislocalize ZO-1, whereas ΔABCCC CagA that lacked the EPIYA-repeat region consisting of multiple EPIYA segments[18] failed to do so (Fig. 1). Notably, ΔABCCC CagA was localized to the cytoplasm in non-polarized MDCK cells or AGS human gastric epithelial cells (Supplementary Fig. 3)[18], but was associated with the membrane in polarized MDCK cells (Fig. 1b), presumably by the amino-terminal CagA region, as recently reported[9]. The ability to mislocalize ZO-1 that had been lost in ΔABCCC CagA was restored in ΔABCC-s CagA (Fig. 1b). Hence, disorganization of the tight junction requires residues 1009–1086 of ABCCC CagA, but is independent of CagA tyrosine phosphorylation. MDCK cells with ZO-1 mislocalized by CagA exhibited non-polarized distributions of apical marker gp135 and basolateral

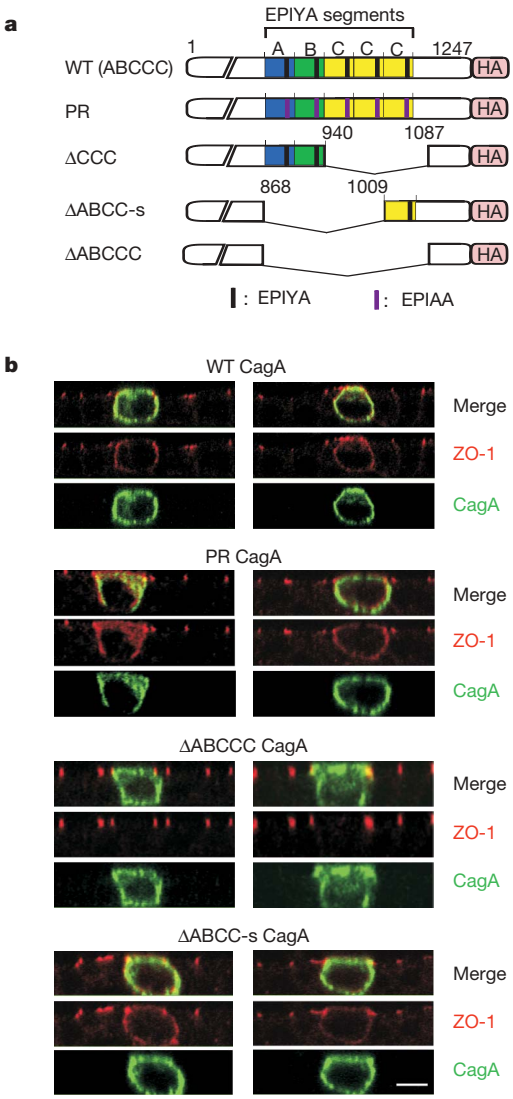


Figure 1 | Disruption of the tight junction by CagA mutants. a, Schematics of CagA constructs. PR, phosphorylation-resistant; WT, wild type. HA, haemagglutinin epitope tag. **b**, Confocal *x*–*z* plane views of ZO-1 localization in MDCK cells expressing various CagA constructs. Scale bar, 10 μm.

[1]Division of Molecular Oncology, Institute for Genetic Medicine and Division of Chemistry, Graduate School of Science, Hokkaido University, Sapporo 060-0815, Japan. [2]Division of Molecular Life Science, Hokkaido University, Graduate School of Life Science, Sapporo 001-0021, Japan. [3]Internal Center for Medical Research and Treatment, Kobe University Graduate School of Medicine, Kobe 650-0017, Japan. [4]Department of Molecular Biology, Yokohama City University Graduate School of Medical Science, Yokohama 236-0004, Japan.

marker E-cadherin, indicating loss of apical–basolateral polarity (Supplementary Fig. 1c). Consequently, CagA-expressing cells showed a 'bowl-shaped' morphology through diminished interaction with neighbouring cells, and more than 50% of them were extruded from the MDCK monolayer within 24 h after transfection of the CagA-expression vector, without signs of apoptosis (Supplementary Fig. 4 and data not shown).

To investigate the mechanism by which CagA causes junctional and polarity defects, we explored host cell proteins that bind to ΔABCC-s but not ΔABCCC CagA. To this end, cell lysates prepared from AGS cells expressing haemagglutinin (HA)-tagged ΔABCC-s or ΔABCCC CagA were immunoprecipitated with an anti-HA antibody and the immunoprecipitates were subjected to proteomic analysis with liquid chromatography/tandem mass spectrometry (LC/MS/MS). Results of the experiment revealed seven proteins that bound to ΔABCC-s but not ΔABCCC CagA (Supplementary Table 1). The protein that showed the highest score was PAR1b/MARK2 serine/threonine kinase (Supplementary Fig. 5)[19], a member of the partitioning-defective 1 (PAR1)/microtubule affinity-regulating kinase (MARK) family that was first isolated in *Caenorhabditis elegans* as a product of one of the six independent 'partitioning-defective' (*par*) genes, mutations in which led to loss of the anterior–posterior polarity of one-cell embryos[20,21].

Physical association of PAR1b with wild-type, PR or ΔABCC-s CagA but not ΔABCCC or ΔCCC CagA was verified by co-transfection experiments in COS-7 cells (Fig. 2a, and Supplementary Fig. 6a). Hence, CagA–PAR1b interaction requires residues 1009–1086 of ABCCC CagA, which are also required for ZO-1 mislocalization. Interaction of CagA with endogenous PAR1b was demonstrated in lysates prepared from AGS cells transfected with a CagA expression vector or *in vitro* infected with an *H. pylori cagA*-positive strain (Fig. 2b, c). ABD-type East Asian CagA[4], another major form of CagA, also bound PAR1b via residues 922–1012 (Supplementary Fig. 6b). Comparison of the PAR1b-binding regions in

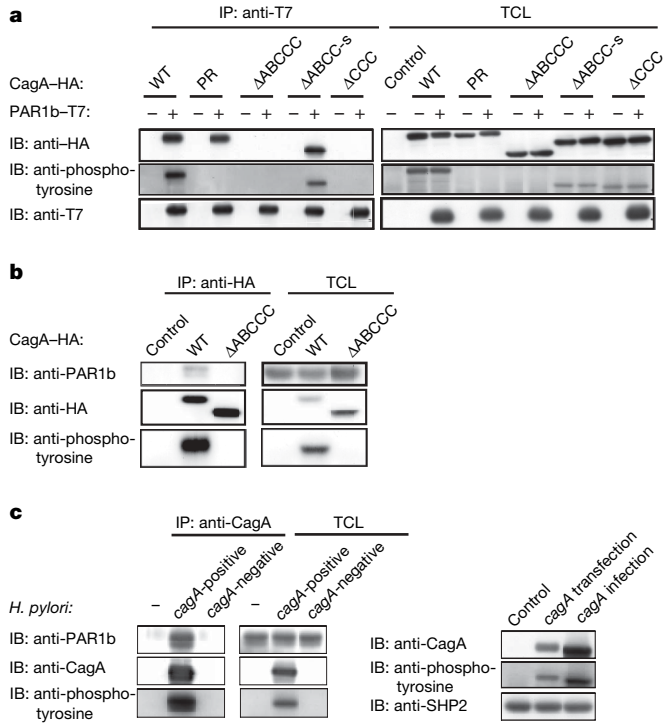


Figure 2 | Interaction of CagA with PAR1. **a**–**c**, COS-7 (**a**) or AGS cells (**b**, **c**) were transfected with the indicated vectors (**a**, **b**) or infected with *H. pylori* NCTC11637 strain or its *cagA*-deleted isogenic strain (**c**). Total cell lysates (TCL) were immunoprecipitated with the respective antibodies and the immunoprecipitates (IP) were immunoblotted (IB) with the indicated antibodies.

Western and East Asian CagA species revealed the presence of a conserved 16-amino-acid sequence, which we previously identified as the CagA-multimerization sequence[15] (Supplementary Fig. 6c). Owing to structural variations within the EPIYA-repeat region, Western CagA proteins possess various numbers of CagA-multimerization sequences, whereas East Asian CagA proteins possess a single CagA-multimerization sequence[15]. Specific deletion of the CagA-multimerization sequences from Western or East Asian CagA abolished the CagA–PAR1b interaction (Supplementary Fig. 6d, e), indicating that the CagA-multimerization sequence constitutes the PAR1 binding site. CagA–PAR1b complex formation was proportional to the number of CagA-multimerization sequences among distinct CagA species (Supplementary Fig. 6f), raising the possibility that the magnitude of individual CagA to elicit junctional and polarity defects is influenced by the number of CagA-multimerization sequences. PAR1b interacted with CagA through the carboxy-terminal region of the kinase catalytic domain (residues 250–276) (Supplementary Fig. 7). The identified CagA-binding sequence was highly conserved among a family of mammalian PAR1 kinases comprising four homologues, PAR1a/MARK3/C-TAK1, PAR1b/MARK2/EMK1, PAR1c/MARK1 and PAR1d/MARK4/MARKL1, which may have redundant functions, and, indeed, all of the human PAR1 kinases specifically bound to CagA (Supplementary Fig. 8).

PAR1 kinases have an essential role in the establishment and maintenance of an organized epithelial architecture[10–13,22,23]. Deletion of the CagA-binding sequence (residues 250–276) from PAR1b abolished the kinase activity (Supplementary Fig. 9), indicating functional importance of the region. Indeed, PAR1b kinase activity was markedly suppressed by binding of CagA in an *in vitro* kinase assay (Fig. 3a). CagA also reduced the phosphorylation of tau protein by PAR1 at Ser 262 (Fig. 3b). Thus, CagA binds PAR1 kinases and inhibits their activities. Expression of a dominant-negative PAR1b (PAR1b MC) induced mislocalization of E-cadherin[22] and ZO-1 in polarized MDCK cells (Supplementary Fig. 10a), which was concomitantly associated with extrusion of bowl-shaped cells from the monolayer, indicating that inhibition of PAR1 phenocopies the CagA activity. We therefore investigated whether elevated levels of PAR1 could oppose the CagA activity in MDCK cells. We produced a recombinant adenovirus that expresses HA-tagged CagA or T7-tagged PAR1b. Infection of the polarized MDCK monolayer with CagA adenovirus induced extrusion of cells from the monolayer, which was abolished by co-infection with PAR1b adenovirus (Fig. 3c). Infection with CagA adenovirus also reduced transepithelial electrical resistance (TER) of the MDCK monolayer, indicating that CagA impaired the tight junction function, which was again reverted by co-infection with PAR1b adenovirus (Fig. 3d). Reciprocally, deletion of the CagA-multimerization sequences, with which PAR1 binds, from CagA abolished the ability of CagA to mislocalize ZO-1 in polarized MDCK cells (Supplementary Fig. 10b). These results collectively indicate that CagA causes junctional and polarity defects through PAR1. Because mammalian PAR1 was initially identified as a MARK kinase that phosphorylates microtubule-associated proteins[19], CagA may inhibit PAR1-dependent microtubule-associated protein phosphorylation and thereby elicit junctional and polarity defects through impaired microtubule-based transport.

In contrast to its active role in basolateral membrane polarization, PAR1 needs to be excluded from the tight junction to maintain apical membrane polarity. This task is achieved by atypical protein kinase C (aPKC), which binds and phosphorylates PAR1 (Thr 595 in the case of PAR1b) at the tight junction[12,13]. The aPKC-mediated phosphorylation induces dissociation of PAR1 from the membrane and thereby prevents the invasion of PAR1 to the apical membrane. Expression of CagA in polarized MDCK cells, however, caused mislocalization of PAR1 to the apical membrane (Fig. 4a). CagA expression also reduced Thr 595 PAR1b phosphorylation and abolished dissociation of PAR1b from the membrane in AGS cells (Fig. 4b, and Supplementary Fig. 11a, b). When ectopically expressed in

COS-7 cells, CagA, PAR1b and aPKCζ were found in a protein complex (Supplementary Fig. 11c), in which PAR1b was phosphorylated on Thr 208 but not on Thr 595 (Supplementary Fig. 11d, lanes 1 and 2). In contrast, PAR1b was phosphorylated on both Thr 208 and Thr 595 in total cell lysates (lanes 6 and 7). Thus, aPKC associates with the CagA–PAR1b complex but fails to phosphorylate CagA-associated PAR1b, causing aberrant localization of PAR1, which may also contribute to polarity defects.

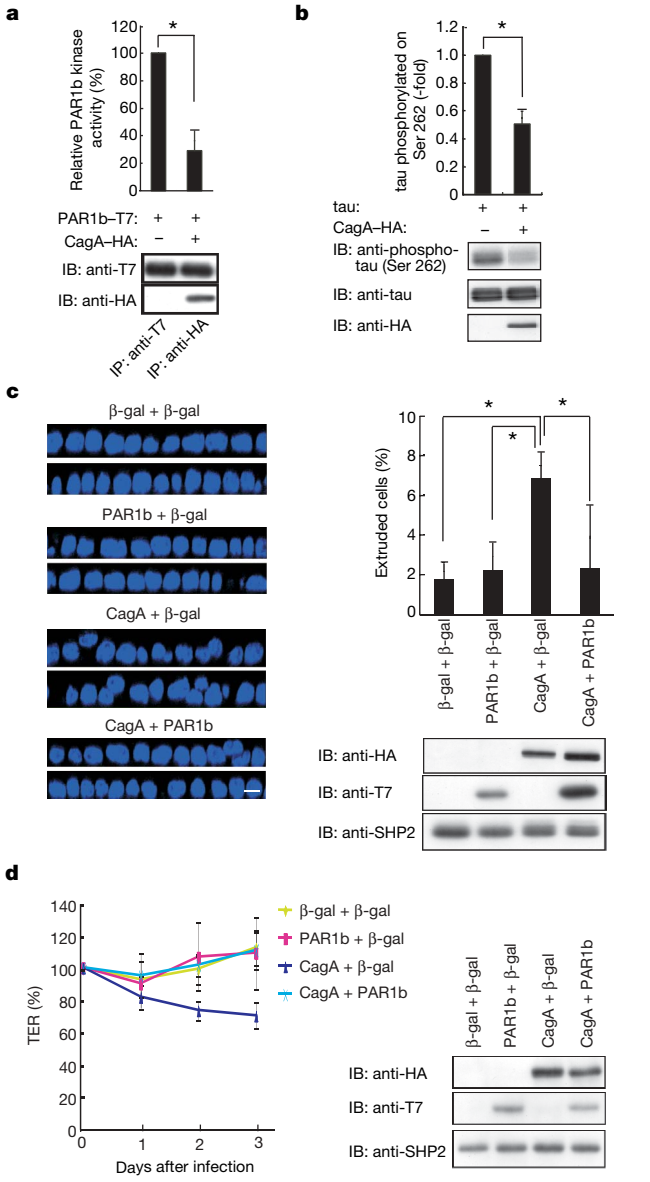


Figure 3 | Inhibition of PAR1 kinase by CagA. **a**, PAR1b or CagA-associated PAR1b immunoprecipitated from COS-7 cells was subjected to an *in vitro* kinase assay. Kinase activity is expressed as the per cent ratio of the tau peptide TR1 phosphorylation in the absence versus presence of CagA. **b**, Lysates of AGS cells transfected with a tau vector with or without wild-type CagA–HA vector were immunoblotted with the indicated antibodies. Quantification of Ser 262 phosphorylation is expressed as the ratio of phosphorylated tau to total tau. **c**, Per cent ratio of cells extruded from the MDCK monolayer (left panel, confocal *x–z* plane views of 4,6-diamidino-2-phenylindole (DAPI) staining) to total MDCK cells at 24 h after co-infection with indicated adenoviruses (multiplicity of infection = 240 for each virus) in a field of 105.86 × 105.86 μm^2 (right, upper panel). Expressed protein levels are shown (right, lower panel). β-gal, β-galactosidase negative control. **d**, TER variation of polarized MDCK monolayer after co-infection with indicated adenoviruses (left panel). Expressed protein levels are shown (right panel). **a–d**, Error bars, ±s.d. ($n = 3$); **a–c**, $^*P < 0.01$, Student's *t*-test.

CagA induces dramatic cell elongation referred to as the hummingbird phenotype[2]. The hummingbird phenotype is caused by the specific interaction of tyrosine-phosphorylated CagA with SHP2 (refs 3–5). Association of CagA with Grb2, c-Met, and/or Crk may also contribute to the morphological change[24–26]. In the present study, we found that CagA binds PAR1 through the CagA-multimerization sequence, which is required for CagA multimerization and subsequent stable association of CagA with SHP2 (ref. 15). The notion that PAR1 exists as a homomultimer (most probably a homodimer)[14] raised the possibility that CagA forced to multimerize through its interaction with a PAR1 multimer. Indeed, CagA, PAR1b and SHP2 were found in a protein complex when expressed in COS-7 cells (Supplementary Fig. 12a), and increased CagA–PAR1 interaction potentiated both CagA multimerization and CagA–SHP2 complex formation (Supplementary Fig. 12b). Given this, we investigated the effect of PAR1b on induction of the hummingbird phenotype by CagA. Despite its stimulatory role in CagA–SHP2 interaction, co-expression of PAR1b suppressed induction of the hummingbird phenotype by CagA (Fig. 4c). A T208A PAR1b mutant[27] that has reduced kinase activity (~10% of WT PAR1b) was significantly less effective in inhibition of the CagA-mediated hummingbird phenotype (Fig. 4c). Thus, active PAR1 kinases suppress the morphogenetic activity of CagA. This in turn indicates that induction of the hummingbird phenotype by CagA requires simultaneous inhibition of PAR1 kinase activity by CagA. Hence, the CagA–PAR1 interaction not only elicits junctional and polarity

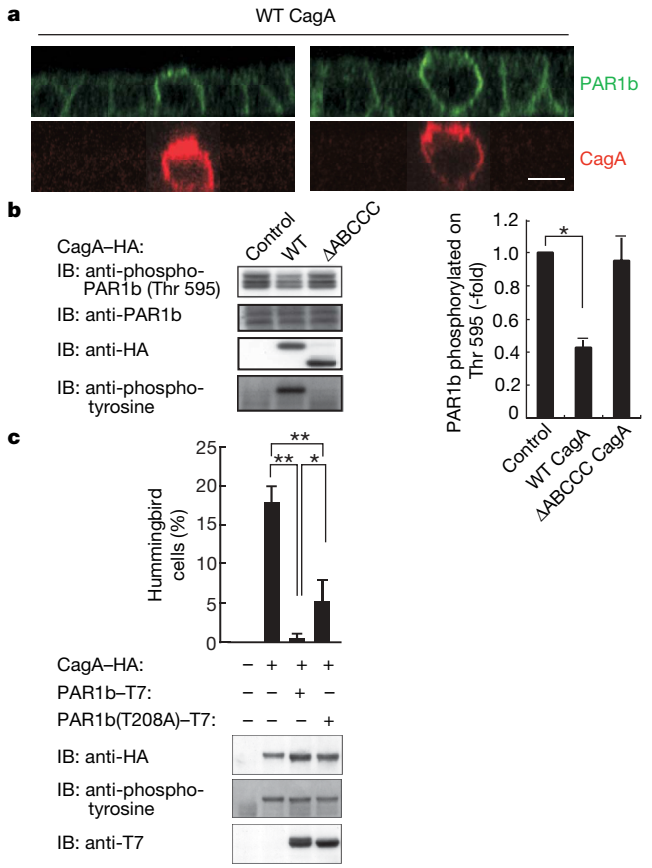


Figure 4 | Role of PAR1 in the biological activities of CagA. **a**, Confocal *x–z* plane views of PAR1b localization in MDCK cells expressing CagA. Scale bar, 10 μm. **b**, AGS cells transfected with the indicated vector were lysed and immunoblotted (IB) with the respective antibodies. Quantification of Thr 595 phosphorylation is expressed as the ratio of phosphorylated PAR1b to total PAR1b. Error bars, ±s.d. ($n = 3$); $^*P < 0.001$, Student's *t*-test. **c**, AGS cells were co-transfected with CagA and indicated PAR1b vectors, and the number of hummingbird cells was counted. Error bars, ±s.d. ($n = 3$); $^*P < 0.01$, $^{**}P < 0.001$, Student's *t*-test.

defects but also has an important role in the morphogenetic activity of CagA (Supplementary Fig. 13).

The present work provides the first molecular link between a human pathogen and the PAR polarity-regulating system that is conserved throughout metazoans[10,11]. On delivery into gastric epithelial cells, *H. pylori* CagA targets PAR1 to disorganize gastric mucosal architecture underlying tissue damage (ulceration) and inflammation (gastritis) *in vivo*. Because normal epithelial architecture constrains cell proliferation[28], its disorganization by CagA may liberate cells from growth-inhibitory cues and thereby promote gastric carcinogenesis in conjunction with the phosphorylation-dependent CagA activities[5]. The human orthologue of *C. elegans par-4* encodes LKB1, which phosphorylates and activates PAR1 (ref. 29). Given that LKB1 is inactivated in the cancer-prone Peutz-Jeghers syndrome[30], our findings also suggest a possible general role for PAR1 inhibition in gastrointestinal carcinogenesis.

METHODS SUMMARY

Cells

MDCK II, COS-7 and AGS cells were used in this work.

Expression vectors. CagA and PAR1b vectors were described previously[12,18,27]. The complementary DNAs encoding PAR1b deletion mutants were cloned into the pEF vector. Human PAR1a, PAR1c and PAR1d vectors were from Origene. The aPKCζ and tau expression vectors were provided by J. Blenis and A. Harada, respectively. Plasmids were transfected into cells by using Lipofectamine 2000 reagent (Invitrogen) or by the calcium phosphate method as previously described[4]. A recombinant adenovirus that expresses CagA, PAR1b or β-galactosidase was generated using ViraPower Adenoviral Expression System (Invitrogen).

Antibodies, immunoprecipitation and immunoblotting. Antibodies used in the work are described in Methods. Immunoprecipitation and immunoblotting were performed as described previously[3,4]. Intensities of chemiluminescence on the immunoblotted filter were quantified by a luminescence-image analyser.

***In vitro* kinase assay.** PAR1b kinase activity was measured via phosphorylation of the tau peptide TR1 at a final concentration of 100 μM.

Mass spectrometry. Immunoprecipitated samples were separated on SDS–polyacrylamide gel electrophoresis (PAGE) gel. Tryptic digested peptides from each gel slice were separated by reverse phase chromatography and directly applied to an ion trap mass spectrometer. The primary ion spectrum data generated by LC/MS/MS were screened against the International Protein Index human and mouse protein databases.

Confocal microscopy. Polarized MDCK cells on a Transwell filter were transfected with an expression vector or infected with recombinant adenoviruses. After incubation, cells were fixed with paraformaldehyde and permeabilized with Triton X-100. The cells were then treated with a primary antibody and were visualized with Alexa Fluor conjugate secondary antibodies (Invitrogen). Images were acquired using a confocal microscope system (OLYMPUS).

Transepithelial electrical resistance. The resistance of the polarized MDCK monolayer was measured by using an electrical resistance system (Millipore).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 February; accepted 21 March 2007.

1. Peek, R. M. Jr & Blaser, M. J. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nature Rev. Cancer* **2,** 28–37 (2002).
2. Segal, E. D., Cha, J., Lo, J., Falkow, S. & Tompkins, L. S. Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proc. Natl Acad. Sci. USA* **96,** 14559–14564 (1999).
3. Higashi, H. *et al.* SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* CagA protein. *Science* **295,** 683–686 (2002).
4. Higashi, H. *et al.* Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc. Natl Acad. Sci. USA* **99,** 14428–14433 (2002).
5. Hatakeyama, M. Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. *Nature Rev. Cancer* **4,** 688–694 (2004).
6. Tartaglia, M. *et al.* Somatic mutations in *PTPN11* in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nature Genet.* **34,** 148–150 (2003).
7. Bentires-Alj, M. *et al.* Activating mutations of the Noonan syndrome-associated *SHP2/PTPN11* gene in human solid tumors and adult acute myelogenous leukemia. *Cancer Res.* **64,** 8816–8820 (2004).
8. Amieva, M. R. *et al.* Disruption of the epithelial apical–junctional complex by *Helicobacter pylori* CagA. *Science* **300,** 1430–1434 (2003).
9. Bagnoli, F., Buti, L., Tompkins, L., Covacci, A. & Amieva, M. R. *Helicobacter pylori* CagA induces a transition from polarized to invasive phenotypes in MDCK cells. *Proc. Natl Acad. Sci. USA* **102,** 16339–16344 (2005).
10. Macara, I. G. Parsing the polarity code. *Nature Rev. Mol. Cell Biol.* **5,** 220–231 (2004).
11. Suzuki, A. & Ohno, S. The PAR–aPKC system: lessons in polarity. *J. Cell Sci.* **119,** 979–987 (2006).
12. Suzuki, A. *et al.* aPKC acts upstream of PAR-1b in both the establishment and maintenance of mammalian epithelial polarity. *Curr. Biol.* **14,** 1425–1435 (2004).
13. Hurov, J. B., Watkins, J. L. & Piwnica-Worms, H. Atypical PKC phosphorylates PAR-1 kinases to regulate localization and activity. *Curr. Biol.* **14,** 736–741 (2004).
14. Panneerselvam, S., Marx, A., Mandelkow, E. M. & Mandelkow, E. Structure of the catalytic and ubiquitin-associated domains of the protein kinase MARK/PAR-1. *Structure* **14,** 173–183 (2006).
15. Ren, S., Higashi, H., Lu, H., Azuma, T. & Hatakeyama, M. Structural basis and functional consequence of *Helicobacter pylori* CagA multimerization in cells. *J. Biol. Chem.* **281,** 32344–32352 (2006).
16. Shin, K., Fogg, V. C. & Margolis, B. Tight junctions and cell polarity. *Annu. Rev. Cell Dev. Biol.* **22,** 207–235 (2006).
17. Noach, L. A., Rolf, T. M. & Tytgat, G. N. Electron microscopic study of association between *Helicobacter pylori* and gastric and duodenal mucosa. *J. Clin. Pathol.* **47,** 699–704 (1994).
18. Higashi, H. *et al.* EPIYA motif is a membrane-targeting signal of *Helicobacter pylori* virulence factor CagA in mammalian cells. *J. Biol. Chem.* **280,** 23130–23137 (2005).
19. Drewes, G. *et al.* MARK, a novel family of protein kinases that phosphorylate microtubule-associated proteins and trigger microtubule disruption. *Cell* **89,** 297–308 (1997).
20. Kemphues, K. J., Priess, J. R., Morton, D. G. & Cheng, N. S. Identification of genes required for cytoplasmic localization in early *C. elegans* embryos. *Cell* **52,** 311–320 (1988).
21. Watts, J. L. *et al. par-6*, a gene involved in the establishment of asymmetry in early *C. elegans* embryos, mediates the asymmetric localization of PAR-3. *Development* **122,** 3133–3140 (1996).
22. Bohm, H., Brinkmann, V., Drab, M., Henske, A. & Kurzchalia, T. V. Mammalian homologues of *C. elegans* PAR-1 are asymmetrically localized in epithelial cells and may influence their polarity. *Curr. Biol.* **7,** 603–606 (1997).
23. Cohen, D., Brennwald, P. J., Rodriguez-Boulan, E. & Müsch, A. Mammalian PAR-1 determines epithelial lumen polarity by organizing the microtubule cytoskeleton. *J. Cell Biol.* **164,** 717–727 (2004).
24. Mimuro, H. *et al.* Grb2 is a key mediator of *Helicobacter pylori* CagA protein activities. *Mol. Cell* **10,** 745–755 (2002).
25. Churin, Y. *et al. Helicobacter pylori* CagA protein targets the c-Met receptor and enhances the motogenic response. *J. Cell Biol.* **161,** 249–255 (2003).
26. Suzuki, M. *et al.* Interaction of CagA with Crk plays an important role in *Helicobacter pylori*-induced loss of gastric epithelial cell adhesion. *J. Exp. Med.* **202,** 1235–1247 (2005).
27. Timm, T. *et al.* MARKK, a Ste20-like kinase, activates the polarity-inducing kinase MARK/PAR-1. *EMBO J.* **22,** 5090–5101 (2003).
28. Bilder, D. Epithelial polarity and proliferation control: links from the *Drosophila* neoplastic tumor suppressors. *Genes Dev.* **18,** 1909–1925 (2004).
29. Lizcano, J. M. *et al.* LKB1 is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1. *EMBO J.* **23,** 833–843 (2004).
30. Jenne, D. E. *et al.* Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nature Genet.* **18,** 38–43 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Blenis, A. Toker, A. Harada and G. Drewes for cDNAs, G. K. Ojakian for antibody and K. Kikuchi for help. I.S. was supported by the Iranian Ministry of Science. This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan and by a research grant from Takeda Science Foundation (M.H.).

Author Contributions I.S., H.H., C.O., M.U., N.M-K., Y.S., H.L., N.O. and T.A. performed biochemical and cell biological experiments and analysed data; T.A., A.S. and S.O. generated reagents and contributed to experimental design; and I.S. and M.H. designed experiments and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.H. (mhata@igm.hokudai.ac.jp).

METHODS

Cells. MDCK II and COS-7 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS). AGS cells were cultured in RPMI 1640 medium supplemented with 10% FBS.

Expression vectors. Expression vectors for WT CagA and its derivatives were described previously[18]. Expression vectors for human PAR1b and the T208A mutant were described previously[12,27]. The cDNAs encoding PAR1b deletion mutants were made from *PAR1b* cDNA by restriction enzyme digestion, site-directed mutagenesis or PCR amplification and were cloned into the pEF vector. Expression vectors for human PAR1a, PAR1c and PAR1d were purchased from Origene. An expression vector for Flag-tagged rat aPKCζ was provided by J. Blenis (Harvard Medical School). An expression vector for human tau protein was provided by A. Harada (University of Tokyo). Expression vectors were transfected into MDCK or AGS cells by using Lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's instructions or into COS-7 cells by the calcium phosphate method as previously described[4]. A recombinant adenovirus that expresses CagA, PAR1b or β-galactosidase was generated according to the instructions of the manufacturer (ViraPower Adenoviral Expression System, Invitrogen).

Antibodies, immunoprecipitation and immunoblotting. Anti-haemagglutinin (HA) (3F10) (Roche), anti-HA (Y-11) and anti-SHP2 (Santa Cruz), anti-T7 (M-21) (Santa Cruz), anti-Myc (9E10), anti-Flag (M2) (Sigma-Aldrich), anti-phosphotyrosine (4G10) (Upstate), anti-PAR1a (Upstate), anti-phospho PAR1b (Thr 595) (Rockland), anti-phospho PAR1b (Thr 208) (Cell Signalling), anti-tau (Calbiochem), anti-phospho tau (Ser 262) (Calbiochem), anti-MARK4 (Cell Signalling), and anti-PAR1b (ref.12) antibodies were used as primary antibodies for immunoblotting, immunoprecipitation and immunostaining. Anti-ZO-1 (Zymed), anti-gp135 (mAb3F2; provided by G. Ojakian), anti-E-cadherin (Transduction Laboratories) and anti-*H. pylori* CagA (AUSTRAL Biologicals) antibodies were used for immunostaining. Immunoprecipitation and immunoblotting were performed as described previously[3,4]. Intensities of chemiluminescence on the immunoblotted filter were quantified with the use of LAS-1000 lumino-image analyser (FUJIFILM) under the condition in which the intensity of the image obtained was proportional to the light intensity.

***In vitro* kinase assay.** PAR1b kinase activity was measured via phosphorylation of the tau peptide TR1 (NVKSKIGSTENLK) at a final concentration of 100 μM in 50 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 2 mM EGTA, 0.5 mM PMSF, 1 mM DTT, 0.5 mM bezamidine and 5 μCi of [γ-^{32}P]-ATP for 2 h at 30 °C. Reactions were stopped by addition of phosphoric acid, and the reaction mixture was applied onto P81 phosphocellulose paper, washed with 100 mM phosphoric acid, dried, and the radioactivity measured using a scintillation counter (Beckman LS 6500).

Mass spectrometry. Immunoprecipitated samples were separated on a 10% SDS–PAGE gel, and the region of the gel containing proteins from about 200 to 50 KDa was cut at intervals of 1–4 mm. Proteins in the gel slices were subjected to reduction, alkylation and tryptic digestion with modified trypsin (Roche) at 37 °C for 14 h. After in-gel digestion, the product peptides were extracted with 5% formic acid and acetonitrile, dried under vacuum, and dissolved in 5% formic acid. Multiply digested peptides from each gel slice were separated by micro-capillary C18-reverse phase chromatography (Michrom BioResources) and directly applied to an LCQ Advantage quadrupole ion trap mass spectrometer (Finnigan). The primary ion spectrum data generated by LC/MS/MS were screened against International Protein Index human and mouse protein databases with the Mascot program (Matrix Science) to identify high scoring proteins.

Confocal microscopy. MDCK cells were seeded on a Transwell filter at 3×10^5 cells cm^{-2}, and polarized MDCK cells were transfected with an expression vector or infected with recombinant adenoviruses. After incubation, cells were fixed with 0.5% paraformaldehyde and permeabilized with 0.5% Triton X-100. The cells were then treated with a primary antibody and were visualized with Alexa Fluor conjugate secondary antibodies (Invitrogen).

Transepithelial electrical resistance. Values were obtained by subtracting the blank values from the filters and the medium.

Digital images. Photo images of immunostained cells were scanned and processed by FLUOVIEW (Olympus). Other pictures and illustrations were prepared by Canvas (ACD system).

LETTERS

Alp7/TACC is a crucial target in Ran-GTPase-dependent spindle formation in fission yeast

Masamitsu Sato[1]† & Takashi Toda[1]

Microtubules are essential intracellular structures involved in several cellular phenomena, including polarity establishment and chromosome segregation[1]. Because the nuclear envelope persists during mitosis (closed mitosis) in fission yeast (*Schizosaccharomyces pombe*), cytoplasmic microtubules must be reorganized into the spindle in the compartmentalized nucleus on mitotic entry[2]. An ideal mechanism might be to take advantage of an evolutionarily conserved microtubule formation system that uses the Ran-GTPase nuclear transport machinery[3–5], but no targets of Ran for spindle formation have been identified in yeast. Here we show that a microtubule-associated protein, Alp7, which forms a complex with Alp14, is a target of Ran in yeast for spindle formation. The Ran-deficient *pim1* mutant (*pim1-F201S*) failed to show mitosis-specific nuclear accumulation of Alp7. Moreover, this mutant exhibited compromised spindle formation and early mitotic delay. Importantly, these defects were suppressed by Alp7 that was artificially targeted to the nucleus by a Ran-independent and importin-α-mediated system. Thus, Ran targets Alp7–Alp14 to achieve nuclear spindle formation, and might differentiate its targets depending on whether the organism undergoes closed or open mitosis.

The fission yeast *pim1*/RanGEF (GTP exchange factor) and *spi1*/Ran mutants show mitotic defects with abnormal microtubule behaviour[6–8]. To understand more precisely the role of the Ran pathway in microtubule formation, we observed the behaviour of *atb2*$^{+}$, which encodes α-tubulin, fused to green fluorescent protein (GFP–Atb2) during mitotic spindle formation in the *pim1* mutant. At 36 °C, GFP–Atb2 in wild-type cells showed normal bipolar spindle structures with a bright GFP–Atb2 signal between the two spindle pole bodies (SPBs, a centrosome equivalent in yeast; Fig. 1a). In contrast, GFP–Atb2 in a *pim1* mutant (*pim1-F201S* allele[6]; see Supplementary Note 1 and Supplementary Fig. 2) had an abnormal spindle structure, which was typically represented by the absence of the intense GFP–Atb2 signal between the two SPBs or of the star-shaped array of microtubules that extended outward from the SPB-to-SPB axis (in total 67%; 20 out of 30 pre-anaphase cells; Fig. 1a, b). The distance between the two SPBs in the *pim1-F201S* mutant was generally shorter than that in the wild-type (see Fig. 2d), indicating that the absence of a proper spindle structure might cause mitotic delay in the *pim1-F201S* mutant. In fact, the frequency of pre-anaphase cells was also elevated in the *pim1-F201S* mutant (Fig. 1c). The *pim1-F201S* mutant used in this study was sensitive to the microtubule-depolymerizing drug TBZ (thiabendazole, Fig. 1d), as previously reported for eight other alleles[8]. These observations confirm that the coupling of Ran- and importin-dependent nuclear transport machinery with spindle formation is evolutionarily conserved in fission yeast, and possibly involves an increase in the nuclear concentration of a certain molecule(s) that is the target of Ran.

In *Xenopus laevis*, several molecules such as TPX2 and NuMA have the capacity to induce microtubule formation when they are unloaded from the importin complex by a gradient of GTP-bound Ran, which is generated by RanGEF in the vicinity of chromosomes[9–11] (see Supplementary Fig. 1 for the schematic). TPX2 and NuMA do not, however, seem to have yeast orthologues[3]. Yeast Ran might thus use distinct molecules as its direct targets for microtubule formation,

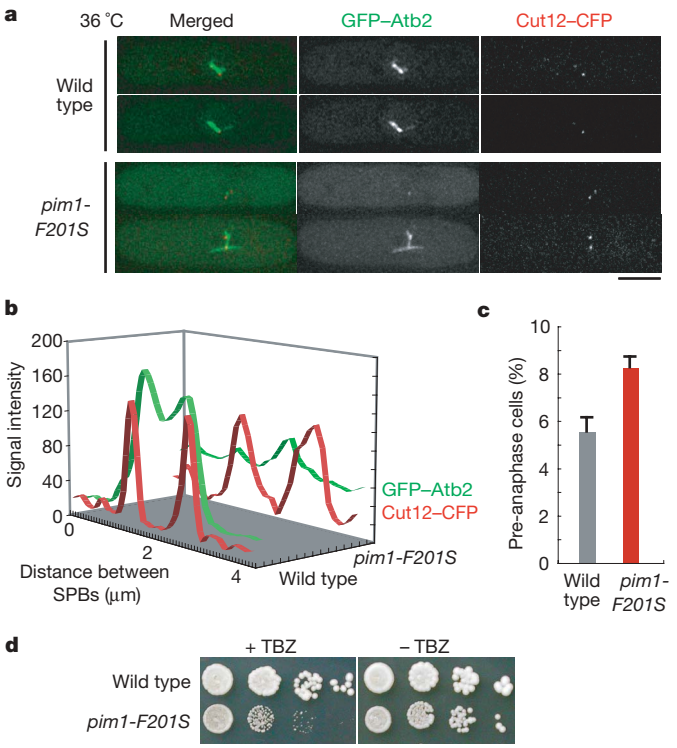


Figure 1 | Defects in bipolar spindle formation in the *pim1* mutant. a, **b**, Cells expressing GFP–Atb2 (α-tubulin) and Cut12–CFP (an SPB marker) in wild-type and *pim1-F201S* backgrounds were inoculated at 36 °C. Scale bar, 5 μm. **b**, GFP–Atb2 and Cut12–CFP signal intensity of a representative cell from each strain was quantified along the inter-SPB line. **c**, Frequency of pre-anaphase cells in wild-type and *pim1-F201S* mutant backgrounds ($n > 270$; error bars, s.d.). **d**, The *pim1-F201S* mutant showed sensitivity to TBZ. Tenfold serial dilutions of the indicated strains were spotted onto medium with or without TBZ (15 μg ml^{-1}).

[1]Laboratory of Cell Regulation, Cancer Research UK, London Research Institute, Lincoln's Inn Fields Laboratories, 44 Lincoln's Inn Fields, London WC2A 3PX, UK. †Present address: Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Hongo, Tokyo 113-0033, Japan.

which presumably would act as microtubule-associated proteins (MAPs). We especially focused on Alp7/Mia1/TACC (transforming acidic coiled-coil protein)[12,13] and its binding partner Alp14/Mtc1/TOG/XMAP215 (tumour overexpressed gene)[14,15] for the following reasons: (i) these MAPs exhibit distinctive nuclear accumulation in addition to localization at nuclear SPBs and/or spindle[12,13] (Fig. 2a; Supplementary Figs 4b and 6); (ii) the mitotic delay seen in the *pim1-F201S* mutant with short inter-SPB distances and fragile spindles is reminiscent of *alp7/alp14* mutants[13,14,16]; and (iii) XMAP215, the *Xenopus* orthologue of Alp14, is a component of the protein complex containing HURP (hepatoma upregulated protein)[17], which was isolated as one of the targets of Ran[17,18]. Because Alp14 failed to accumulate in the nucleus in the *alp7*-deleted (*alp7*Δ) strain, but the converse was not true[13] (Supplementary Fig. 4b), the upstream Alp7, rather than Alp14 might be the binding partner of the importin complex. We thus performed systematic deletion analyses of Alp7 to identify the domain responsible for its nuclear localization; we found a canonical classical nuclear localization signal (NLS) sequence that is rich in basic amino acids (Supplementary Figs 4 and 5). The R122A R124A mutations of Alp7 fused to three tandem copies of GFP (Alp7-RARA–3GFP) resulted in a loss of both Alp7 nuclear localization and function, suggesting that the nuclear import of Alp7, as driven by its intrinsic NLS, is essential for its function in the nucleus (Fig. 2b, c). Importantly, in the *pim1-F201S* mutant, 67% of pre-anaphase cells did not accumulate Alp7–3GFP in the nucleus ($n = 45$; Fig. 2d, e), which suggests that Alp7 is one of the nuclear import targets of Ran. The early mitotic delay in *pim1-F201S* indicated by short inter-SPB distance (<1 μm) is tightly correlated with a loss of Alp7 nuclear accumulation (Supplementary Fig. 3).

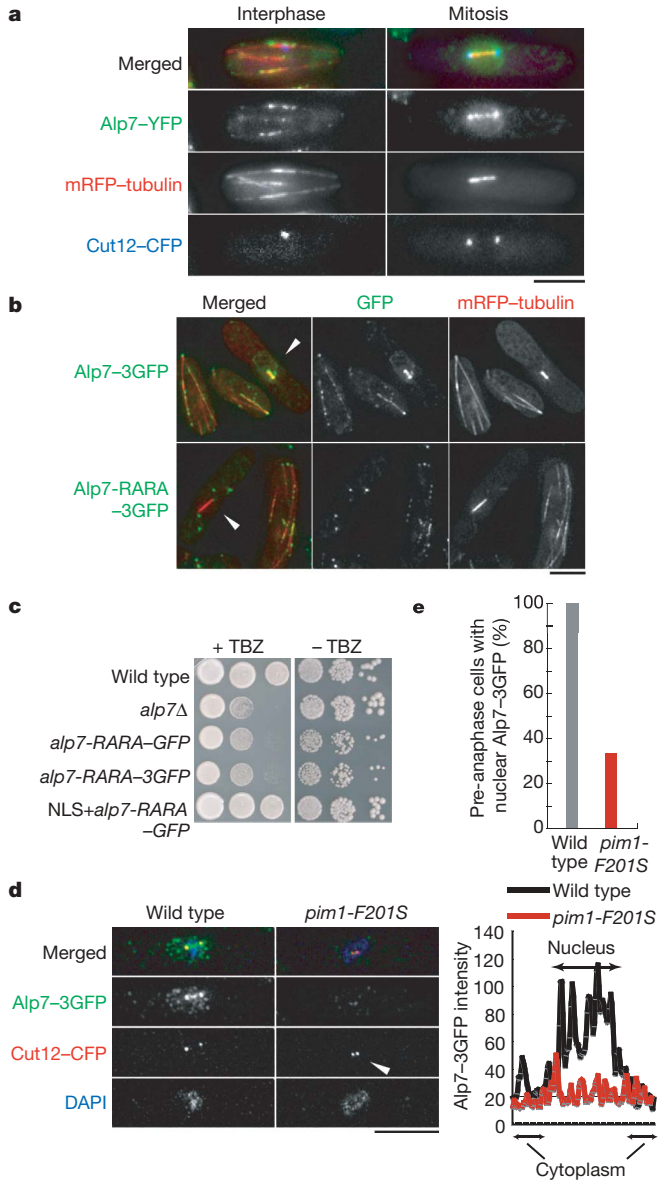


Figure 2 | Alp7 is a target of the Ran-dependent transport machinery. **a**, Cell-cycle-dependent localization of Alp7–YFP. mRFP–tubulin and Cut12–CFP were visualized simultaneously in wild-type yeast. **b**, Alp7–3GFP and NLS-deficient Alp7-RARA–3GFP were each observed with mRFP–tubulin. Arrowheads indicate mitotic nuclei. **c**, The intrinsic NLS of Alp7 is essential for its function. Addition of an external NLS to Alp7-RARA restored its function. **d**, **e**, The *pim1-F201S* mutant does not accumulate Alp7 in mitotic nuclei. **d**, The arrowhead indicates short inter-SPB distance (<1 μm). The right panel shows a representative example of the quantification of Alp7–3GFP signal. See Methods for quantification of fluorescence signal. **e**, Frequency of pre-anaphase cells with nuclear Alp7–3GFP. Scale bars, 5 μm.

As Ran is a general regulator of nuclear transport, it might interact with many 'cargo' molecules involved in microtubule organization. Is Alp7 the key molecule among them? To address this question, we undertook a strategy to accumulate Alp7 artificially in the nucleus in the *pim1* mutant and to determine whether this would suppress the observed mitotic defects. Suppression would indicate that Alp7 is one of Ran's crucial targets. In the *pim1-F201S* mutant, nuclear targeting of Alp7 by NLS-tagging is not, however, effective because of disabled Ran function[6]. An alternative would be to use the Ran-independent transport system, although no precedent for this approach has been reported in fission yeast. In mammals and budding yeast, truncated importin-α molecules, which neither bind to importin-β nor act as a Ran-dependent cargo protein, are still able to enter the nucleus[19]. We therefore constructed a variant of one of the fission yeast importin-α genes, *imp1* (ref. 20), which encodes a protein that lacks the importin-β- and cargo-binding domains (Imp1ΔIBB-ED). GFP–Imp1ΔIBB-ED accumulated in the nucleus (Fig. 3a), confirming that Ran-independent transport of importin-α might also be conserved in fission yeast. Furthermore, the GFP–Imp1ΔIBB-ED–Alp7 fusion product localized in the nucleus, even during interphase, together with Alp14–rRFP (refoldable red fluorescent protein; Fig. 3b), suggesting that Imp1ΔIBB-ED–Alp7 retained the ability to bind Alp14. Unexpectedly, however, Imp1ΔIBB-ED–Alp7 was unable to localize to SPBs or spindles (data not shown). Imp1ΔIBB-ED probably tethered Alp7 to a nuclear site such as the nuclear periphery, and as a result Alp7 could not localize to SPBs.

To circumvent this situation, the TEV protease was also introduced into these cells. This would allow cleavage of Imp1ΔIBB-ED–Alp7 to liberate Alp7 from the fusion after being transported into the nucleus. For this purpose, Imp1ΔIBB-ED and Alp7 were linked by the recognition site for TEV protease (Imp1ΔIBB-ED–tev–Alp7, simply abbreviated 'Imp1*–tev–Alp7' hereafter), and the wild-type *alp7*$^+$ ORF was replaced by the fusion construct to express it endogenously under the *alp7*$^+$ promoter. Simultaneously, Imp1ΔIBB-ED-tagged TEV protease (Imp1*–TEVp; which includes a cyan fluorescent protein, CFP, tag) was expressed. In fact, co-expression of these two Imp1ΔIBB-ED-fusion constructs (denoted as the Imp1*–tev–Alp7 system hereafter; Supplementary Fig. 7) allowed site-specific cleavage of Imp1*–tev–Alp7 in the nucleus (~50% was cleaved *in vivo*; Fig. 3c). As expected, Alp7–3GFP liberated from Imp1ΔIBB-ED localized to SPBs and/or spindle during mitosis or was exported to the cytoplasm during interphase (Fig. 3d and Supplementary Fig. 8). Finally, the Imp1*–tev–Alp7 system restored nuclear accumulation and SPB and/or spindle localization of Alp7 in the Ran-deficient *pim1-F201S* mutant only in the presence of Imp1*–TEVp (Fig. 3e, f), providing genetic evidence to support Ran-independent transport of importin-α.

Having established the Ran-independent nuclear import and unloading system for Alp7, we examined whether the Imp1*–tev–Alp7 system could suppress mitotic defects of the *pim1-F201S*

mutant. Induction of the Imp1*–tev–Alp7 system in the *pim1-F201S* mutant ameliorated the mitotic delay (Fig. 4a) and restored spindle microtubule formation (Fig. 4b). The frequency of pre-anaphase cells with an intense GFP–Atb2 signal increased from 33% to 89% with expression of the Imp1*–tev–Alp7 system ($n = 75$; Fig. 4c). The *pim1-F201S* mutant also showed non-bipolar spindles such as monopolar or star-shaped spindles (Fig. 1a and data not shown). However, 57% of pre-anaphase cells from the *pim1-F201S* mutant with the Imp1*–tev–Alp7 system exhibited bipolar spindles with a bright GFP–Atb2 signal, in contrast to 23% of *pim1-F201S* cells without the Imp1*–tev–Alp7 system ($n = 60$; Fig. 4c). Finally, growth of the *pim1-F201S* mutant at a semi-restrictive temperature (34 °C) was rescued by the Imp1*–tev–Alp7 system, albeit partially (Fig. 4d). Thus, artificial nuclear targeting of Alp7 restored spindle formation and suppressed the mitotic delay and/or growth defect of the *pim1*/RanGEF mutant. These results consistently indicate that Alp7 is one of the essential targets of Ran GTPase for spindle formation. In the absence of Alp14, however, the Imp1*–tev–Alp7 system was unable to restore either temperature sensitivity or spindle defects of the *pim1-F201S* mutant ($n = 46$; Fig. 4c–e). In addition, Alp7–NLS was incapable of suppressing *alp14*Δ mutant (Fig. 4f), indicating that Alp14 is indispensable for Alp7 function in spindle formation. These results imply that Alp7 is a direct target of the Ran-dependent import machinery but its binding partner Alp14 may be the factor that promotes spindle formation once it is loaded onto the mitotic SPB via Alp7.

pim1 mutants exhibit multiple phenotypes such as incomplete cell separation and a defect in chromosome decondensation[6,21,22]. The Imp1*–tev–Alp7 system did not, however, overcome those defects (data not shown; Supplementary Note 2), which indicates that Alp7 is a target of Ran at least at early stages of spindle formation, for example, for increasing local tubulin and/or microtubule concentrations around nuclear SPBs, or for nucleation from the SPBs (Supplementary Fig. 1).

Alp7 had never been identified as a direct target of Ran in other organisms. Why is the Alp7–Alp14 complex a crucial target of Ran in yeast, instead of TPX2, NuMA or HURP as has been shown for higher organisms? In those organisms, for instance, TPX2 activates Aurora A kinase, which in turn phosphorylates TACC at centrosomes[23,24]. Thus, it is possible that this molecular cascade might be bypassed by direct targeting of Alp7–Alp14 by Ran, probably because analogous intermediate proteins or systems do not exist and/or are not necessitated in fission yeast[3] (Supplementary Note 3).

Ran exerts influence on all cellular processes by the same mechanism: RanGTP destabilizes affinity between importins and cargos in the nuclei during interphase. In the open mitosis of higher organisms, nuclear envelope breakdown allows dissociation of the complexes in the immediate vicinity of the chromosomes, resulting in microtubule formation around chromosomes[25]. The diameter of a RanGTP gradient that has the potential to induce microtubule nucleation around a chromosome in *Xenopus* egg extracts is comparable to the size of the yeast nucleus[26,27]. Thus, it may not be possible for an effective local gradient of RanGTP to form inside the yeast nucleus.

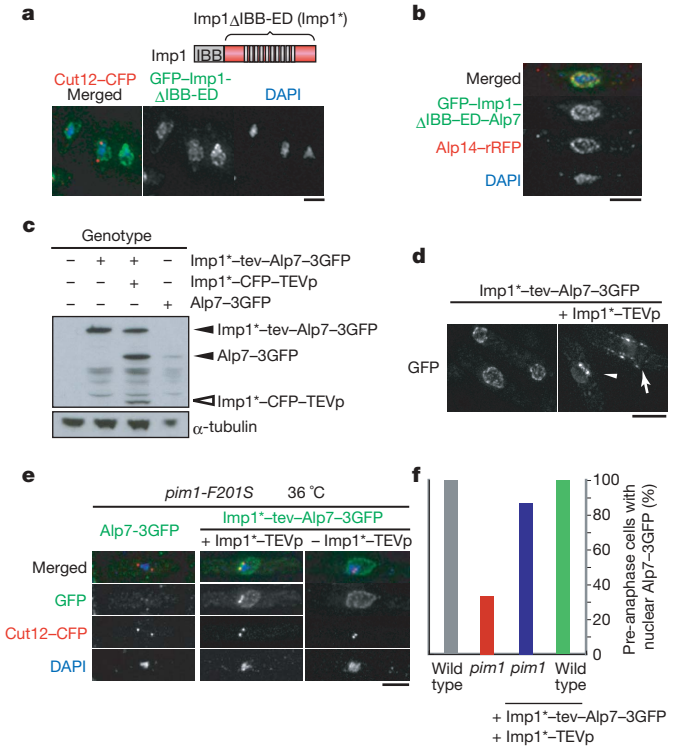


Figure 3 | Establishment of a Ran-independent nuclear import and unloading system by importin-α and TEV protease. **a**, Cut12–CFP cells expressing GFP–Imp1ΔIBB-ED were stained with DAPI. **b**, The GFP–Imp1ΔIBB-ED–Alp7 fusion protein accumulated in the nucleus together with Alp14–rRFP. **c**, *In vivo* cleavage of Imp1*–tev–Alp7–3GFP by Imp1*–TEVp. 'Imp1*' represents Imp1ΔIBB-ED. **d**, Imp1*–tev–Alp7–3GFP cells with (right panel) or without (left panel) Imp1*–CFP–TEVp. Arrow, interphase cell; arrowhead, mitotic cell. **e**, **f**, Mitotic nuclear localization of Alp7 in the *pim1-F201S* mutant was restored by the Imp1*–tev–Alp7 system. The population of pre-anaphase cells with nuclear Alp7–3GFP was also quantified ($n > 186$; **f**). Scale bars, 5 μm.

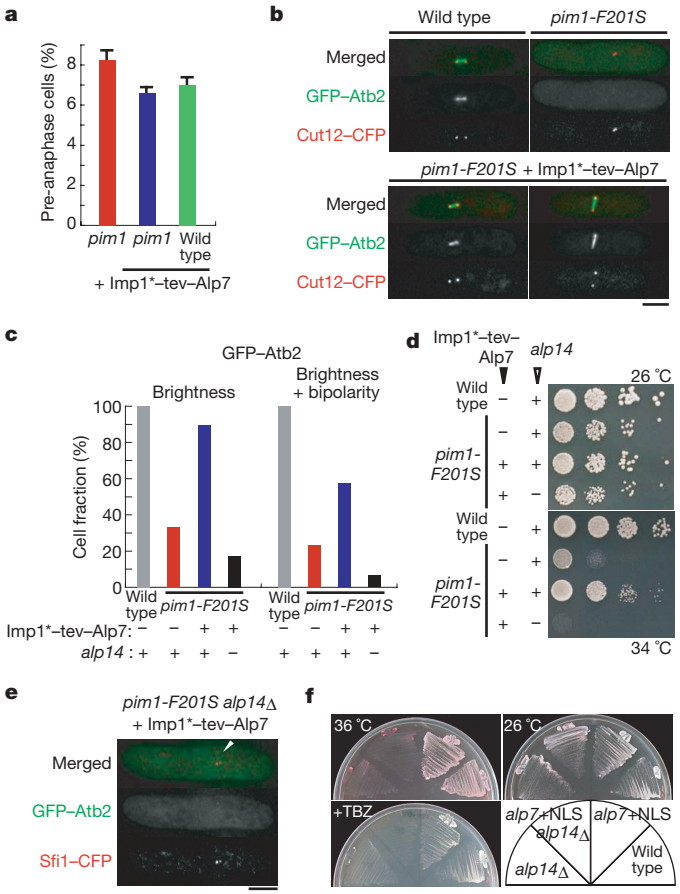


Figure 4 | The Imp1*–tev–Alp7 system restores spindle formation and mitotic progression of the *pim1-F201S* mutant. **a**, The proportion of pre-anaphase cells in the *pim1* mutant with the Imp1*–tev–Alp7 system ($n > 329$; error bars, s. d.). **b**, **c**, GFP–Atb2 in wild-type and *pim1-F201S* backgrounds. **c**, The frequency of pre-anaphase cells with an intense GFP–Atb2 signal in a wild-type (+) or *alp14*Δ (–) background. **d**, Effect of the Imp1*–tev–Alp7 system on the growth of the *pim1-F201S* mutant. **e**, GFP–tubulin in *pim1-F201S alp14*Δ + Imp1*–tev–Alp7 cell (arrowhead: SPB). **f**, The indicated strains were streaked on rich media. Scale bars, 5 μm.

By targeting Alp7, however, this problem might be solved, because Alp7 preferentially localizes to SPBs[13], thereby concentrating nuclear tubulin and/or microtubules or promoting bipolar spindle formation locally around the SPBs or possibly the chromosomes (Supplementary Note 4). For efficient spindle formation in organisms with open mitosis that have larger spindles than do yeast cells, Ran's multiple targets such as TPX2 or NuMA might be assigned to coordinate long-range communication between microtubules and chromatin via MAPs, including TACC–TOG[28]. Thus, Ran might use distinct targets depending on the environment where spindle microtubules are formed.

METHODS SUMMARY

Yeast genetics and strains. Conventional methods for yeast genetics were used. The truncation, fusion and site-directed mutation constructs used in this study were integrated into the original loci of chromosomes to achieve endogenous expression patterns, unless otherwise stated; see Methods. Strains used in this study are listed in Supplementary Tables 1 and 2; see Supplementary Note 1 for details of the *pim1-F201S*/RanGEF mutant[6], which we used.

Microscopy. For fluorescence microscopy, an AxioplanII microscope (Zeiss) with Volocity software (Improvision) and the Deltavision-SoftwoRx system (Olympus and Applied Precision) were used. Images were taken as 6–14 sections along the *z* axis at 0.2–0.4-μm intervals; they were then deconvolved and merged into a single projection. For observation of the *pim1* mutant, cells were grown at 26 °C overnight and then shifted to 36 °C for 4 h. Cells were subsequently observed under the microscope at 36 °C (total duration at 36 °C including observation was ≤5 h) or fixed with formaldehyde and stained with 4,6-diamidino-2-phenylindole (DAPI). See Methods for quantification of fluorescence signal.

Constructs. The detailed construct for Imp1*–tev–Alp7 is illustrated in Supplementary Fig. 7. Briefly, a fission yeast importin-α protein, Imp1[19,20], was cloned without the amino-terminal 52 amino acids (Imp1ΔIBB). Mutations to the putative cargo-binding sites of Imp1 (D192R E393R) were introduced to abolish its cargo-binding ability[9,19]. The resulting Imp1ΔIBB-ED construct is simply denoted as Imp1* in the text. The recognition site for TEV protease was inserted between Imp1ΔIBB-ED and Alp7, and the fusion construct (Imp1*–tev–Alp7) was integrated into the *alp7* locus, followed by GFP- or monomeric red fluorescent protein (mRFP)-tagging. For an Imp1ΔIBB-ED–TEV protease fusion, the Imp1ΔIBB-ED fragment was cloned into the *adh81*-driven CFP–TEV protease plasmid and was integrated into the chromosome[29] (Supplementary Fig. 7).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 December 2006; accepted 23 March 2007.
Published online 2 May 2007.

1. Kirschner, M. & Mitchison, T. Beyond self-assembly: from microtubules to morphogenesis. *Cell* **45,** 329–342 (1986).
2. Hagan, I. M. & Hyams, J. S. The use of cell division cycle mutants to investigate the control of microtubule distribution in the fission yeast *Schizosaccharomyces pombe*. *J. Cell Sci.* **89,** 343–357 (1988).
3. Hetzer, M., Gruss, O. J. & Mattaj, I. W. The Ran GTPase as a marker of chromosome position in spindle formation and nuclear envelope assembly. *Nature Cell Biol.* **4,** E177–E184 (2002).
4. Karsenti, E. & Vernos, I. The mitotic spindle: a self-made machine. *Science* **294,** 543–547 (2001).
5. Zheng, Y. G protein control of microtubule assembly. *Annu. Rev. Cell Dev. Biol.* **20,** 867–894 (2004).
6. Hirose, E. *et al.* Loss of RanGEF/Pim1 activity abolishes the orchestration of Ran-mediated mitotic cellular events in *S. pombe*. *Genes Cells* **11,** 29–46 (2006).
7. Fleig, U., Salus, S. S., Karig, I. & Sazer, S. The fission yeast ran GTPase is required for microtubule integrity. *J. Cell Biol.* **151,** 1101–1111 (2000).
8. Salus, S. S., Demeter, J. & Sazer, S. The Ran GTPase system in fission yeast affects microtubules and cytokinesis in cells that are competent for nucleocytoplasmic protein transport. *Mol. Cell. Biol.* **22,** 8491–8505 (2002).
9. Gruss, O. J. *et al.* Ran induces spindle assembly by reversing the inhibitory effect of importin α on TPX2 activity. *Cell* **104,** 83–93 (2001).
10. Wiese, C. *et al.* Role of importin-β in coupling Ran to downstream targets in microtubule assembly. *Science* **291,** 653–656 (2001).
11. Nachury, M. V. *et al.* Importin β is a mitotic target of the small GTPase Ran in spindle assembly. *Cell* **104,** 95–106 (2001).
12. Oliferenko, S. & Balasubramanian, M. K. Astral microtubules monitor metaphase spindle alignment in fission yeast. *Nature Cell Biol.* **4,** 816–820 (2002).
13. Sato, M., Vardy, L., Garcia, M. A., Koonrugsa, N. & Toda, T. Interdependency of fission yeast Alp14/TOG and coiled coil protein Alp7 in microtubule localization and bipolar spindle formation. *Mol. Biol. Cell* **15,** 1609–1622 (2004).
14. Garcia, M. A., Vardy, L., Koonrugsa, N. & Toda, T. Fission yeast ch-TOG/XMAP215 homologue Alp14 connects mitotic spindles with the kinetochore and is a component of the Mad2-dependent spindle checkpoint. *EMBO J.* **20,** 3389–3401 (2001).
15. Nakaseko, Y., Goshima, G., Morishita, J. & Yanagida, M. M phase-specific kinetochore proteins in fission yeast: microtubule-associating Dis1 and Mtc1 display rapid separation and segregation during anaphase. *Curr. Biol.* **11,** 537–549 (2001).
16. Sato, M. *et al.* Deletion of Mia1/Alp7 activates Mad2-dependent spindle assembly checkpoint in fission yeast. *Nature Cell Biol.* **5,** 764–766 (2003).
17. Koffa, M. D. *et al.* HURP is part of a Ran-dependent complex involved in spindle formation. *Curr. Biol.* **16,** 743–754 (2006).
18. Sillje, H. H., Nagel, S., Korner, R. & Nigg, E. A. HURP is a Ran-importin β-regulated protein that stabilizes kinetochore microtubules in the vicinity of chromosomes. *Curr. Biol.* **16,** 731–742 (2006).
19. Miyamoto, Y. *et al.* Importin α can migrate into the nucleus in an importin β- and Ran-independent manner. *EMBO J.* **21,** 5833–5842 (2002).
20. Umeda, M., Izaddoost, S., Cushman, I., Moore, M. S. & Sazer, S. The fission yeast *Schizosaccharomyces pombe* has two importin-α proteins, Imp1p and Cut15p, which have common and unique functions in nucleocytoplasmic transport and cell cycle progression. *Genetics* **171,** 7–21 (2005).
21. Matsumoto, T. & Beach, D. Premature initiation of mitosis in yeast lacking RCC1 or an interacting GTPase. *Cell* **66,** 347–360 (1991).
22. Sazer, S. & Nurse, P. A fission yeast RCC1-related protein is required for the mitosis to interphase transition. *EMBO J.* **13,** 606–615 (1994).
23. Brittle, A. L. & Ohkura, H. Centrosome maturation: Aurora lights the way to the poles. *Curr. Biol.* **15,** R880–R882 (2005).
24. Gruss, O. J. & Vernos, I. The mechanism of spindle assembly: functions of Ran and its target TPX2. *J. Cell Biol.* **166,** 949–955 (2004).
25. Heald, R. *et al.* Self-organization of microtubules into bipolar spindles around artificial chromosomes in *Xenopus* egg extracts. *Nature* **382,** 420–425 (1996).
26. Kalab, P., Weis, K. & Heald, R. Visualization of a Ran-GTP gradient in interphase and mitotic *Xenopus* egg extracts. *Science* **295,** 2452–2456 (2002).
27. Caudron, M., Bunt, G., Bastiaens, P. & Karsenti, E. Spatial coordination of spindle assembly by chromosome-mediated signaling gradients. *Science* **309,** 1373–1376 (2005).
28. Carazo-Salas, R. E. & Karsenti, E. Long-range communication between chromatin and microtubules in *Xenopus* egg extracts. *Curr. Biol.* **13,** 1728–1733 (2003).
29. Yokobayashi, S. & Watanabe, Y. The kinetochore protein Moa1 enables cohesion-mediated monopolar attachment at meiosis I. *Cell* **123,** 803–817 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Karsenti for comments and R. E. Carazo-Salas and M. Toya for critical reading of the manuscript, discussion and technical support for microscopy. We also thank M. Yoshida, H. Maekawa, E. Schiebel, I. Hagan, R. Y. Tsien, S. Yokobayashi and Y. Watanabe for plasmids and/or information. We are grateful to members of the Cell Regulation Laboratory and M. Yamamoto for support. This work is supported by Cancer Research UK. M.S. is a recipient of a Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship for Research Abroad.

Author Contributions The experiments were designed by M.S with the support of T.T. and performed by M.S. M.S. and T.T. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.S. (masasato@biochem.s.u-tokyo.ac.jp) or T.T. (Takashi.Toda@cancer.org.uk).

METHODS

Strains, gene deletion and tagging. Standard protocols were used for gene tagging and deletion[30,31]. Carboxy-terminal 3GFP-or 4mRFP-tagging into chromosomal genes was performed by a conventional PCR-based method with pMS-3GK (for 3GFP) or pMS-4RK (for 4mRFP) template plasmids, respectively. pMS-3GK was constructed from pSM1023 (ref. 32) by inserting a 20-base oligonucleotide (the sequence to which the forward primer anneals) which has been commonly used in fission yeast gene targeting[30,31] to allow utilization of existing primers designed for other template plasmids. For GFP–Atb2 observation in wild-type and the *pim1-F201S* mutant, we used a strain in which GFP–Atb2 expression is controlled by the $atb2^{+}$ promoter and terminator, because the *nmt1-GFP–Atb2* strain (regulated by the *nmt1* promoter) in our stock showed temperature sensitivity at 36 °C, the restrictive temperature of the *pim1* mutant (data not shown).

Microscopy (signal intensity quantification and mRFP-tubulin expression). GFP–Atb2 and Cut12–CFP signals observed in living cells (Fig. 1b) were quantified with ImageJ software (NIH) as previously described[33]. Briefly, a straight line connecting two SPBs was manually drawn, and the signal intensity along this line was calculated and plotted. 'Brightness' in Fig. 4c was quantified similarly. 'Bipolarity' was evaluated from the existence of SPB signal by Cut12–CFP or Sfi1–CFP at both ends of a spindle. Alp7–3GFP signal intensity in the nuclei of wild-type and the *pim1-F201S* mutant (Fig. 2d, e and Supplementary Fig. 3) was quantified similarly. If the nuclear Alp7–3GFP intensity of a mitotic cell was more than twofold compared with the cytoplasmic signal, we categorized the cell as 'Nuclear Alp7' (for nuclear accumulation) and otherwise as 'Dispersed Alp7' (for no nuclear accumulation) (Supplementary Fig. 3). mRFP–tubulin was expressed from a pREP1 plasmid under the repressive condition (in the presence of thiamine) as described previously[34].

Confirmation of Imp1*–tev–Alp7 cleavage by immunoblotting. To detect *in vivo* cleavage of the fusion construct by TEV protease, protein extracts from individual strains were prepared and subjected to SDS–PAGE, followed by western blotting with anti-GFP (Roche) and anti-tubulin antibodies (Sigma).

30. Bähler, J. *et al.* Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14,** 943–951 (1998).
31. Sato, M., Dhut, S. & Toda, T. New drug-resistant cassettes for gene disruption and epitope tagging in *Schizosaccharomyces pombe*. *Yeast* **22,** 583–591 (2005).
32. Usui, T., Maekawa, H., Pereira, G. & Schiebel, E. The XMAP215 homologue Stu2 at yeast spindle pole bodies regulates microtubule dynamics and anchorage. *EMBO J.* **22,** 4779–4793 (2003).
33. Sato, M. *et al.* Deletion of Mia1/Alp7 activates Mad2-dependent spindle assembly checkpoint in fission yeast. *Nature Cell Biol.* **5,** 764–766 (2003).
34. Yamashita, A., Sato, M., Fujita, A., Yamamoto, M. & Toda, T. The roles of fission yeast Ase1 in mitotic cell division, meiotic nuclear oscillation, and cytokinesis checkpoint signaling. *Mol. Biol. Cell* **16,** 1378–1395 (2005).

Vol 447|17 May 2007|doi:10.1038/nature05720

LETTERS

RNA-templated DNA repair

Francesca Storici[1], Katarzyna Bebenek[1], Thomas A. Kunkel[1], Dmitry A. Gordenin[1] & Michael A. Resnick[1]

RNA can act as a template for DNA synthesis in the reverse transcription of retroviruses and retrotransposons[1] and in the elongation of telomeres[2]. Despite its abundance in the nucleus, there has been no evidence for a direct role of RNA as a template in the repair of any chromosomal DNA lesions, including DNA double-strand breaks (DSBs), which are repaired in most organisms by homologous recombination or by non-homologous end joining[3]. An indirect role for RNA in DNA repair, following reverse transcription and formation of a complementary DNA, has been observed in the non-homologous joining of DSB ends[4,5]. In the yeast *Saccharomyces cerevisiae*, in which homologous recombination is efficient[3], RNA was shown to mediate recombination, but only indirectly through a cDNA intermediate[6,7] generated by the reverse transcriptase function of Ty retrotransposons in Ty particles in the cytoplasm[8]. Although pairing between duplex DNA and single-strand (ss)RNA can occur *in vitro*[9,10] and *in vivo*[11], direct homologous exchange of genetic information between RNA and DNA molecules has not been observed. We show here that RNA can serve as a template for DNA synthesis during repair of a chromosomal DSB in yeast. The repair was accomplished with RNA oligonucleotides complementary to the broken ends. This and the observation that even yeast replicative DNA polymerases such as α and δ can copy short RNA template tracts *in vitro* demonstrate that RNA can transfer genetic information *in vivo* through direct homologous interaction with chromosomal DNA.

We predicted that if RNA could participate directly in the repair of a chromosomal DSB, this would require DNA synthesis on the RNA template. Such activity might be mediated by a reverse transcriptase able to function in the nucleus on the chromosome or possibly by a DNA polymerase. In RNA–protein complexes from LINE1 retrotransposons, retrotranscription can be primed in the nucleus by the 3′ end of a chromosomal break[12]. However, break repair using LINE1 elements does not require RNA/DNA complementarity and is, therefore, mutagenic. On the other hand, it is unclear if DNA polymerases actually possess RNA-templated DNA synthesis activity *in vivo*, despite the ability of *Escherichia coli* Pol I (ref. 13), mammalian Pol γ (ref. 14), and human DNA pol-η, -ι and -κ (ref. 15) to synthesize DNA on RNA templates *in vitro*. Here we explore the possibility that a DSB can be repaired by homologous RNA and that chromosomal DNA synthesis can occur on RNA templates in the yeast *S. cerevisiae*.

We first investigated DNA synthesis across a short RNA tract during the repair of a chromosomal DSB, using the capacity of ssDNA oligonucleotides to serve as a template for efficient DSB repair[16,17]. A site-specific DSB was induced within the *LEU2* gene by overexpression of HO endonuclease[18]. Following DSB induction, cells were transformed with ssDNA oligonucleotides that were designed (see Fig. 1) to join *LEU2* ends and introduce a unique, in-frame 12-base insert containing 0, 4, 6 or 12 RNA bases (*a*, *b*, *c* and *d*, respectively), or a 6-base insert with 0 or 6 RNA bases (*e* and *f*). To accomplish DSB repair and restore a functional *LEU2* gene, the insert sequence must be used as a template. Remarkably, repair by *b* (containing 4 ribonucleotides) was only a factor of 3 lower than repair by the DNA-only control (*a*). The frequencies decreased with increasing RNA-tract length (*b*, *c*, *d*, Fig. 1). The appearance of the oligonucleotide sequence in the Leu$^+$ transformants (see 'Verification' in Supplementary Table 2a) suggests that the RNA-containing molecules participate directly in the repair. This contrasts with the observation that two ribonucleotides at the mating type locus of *Schizosaccharomyces pombe* seem to block DNA

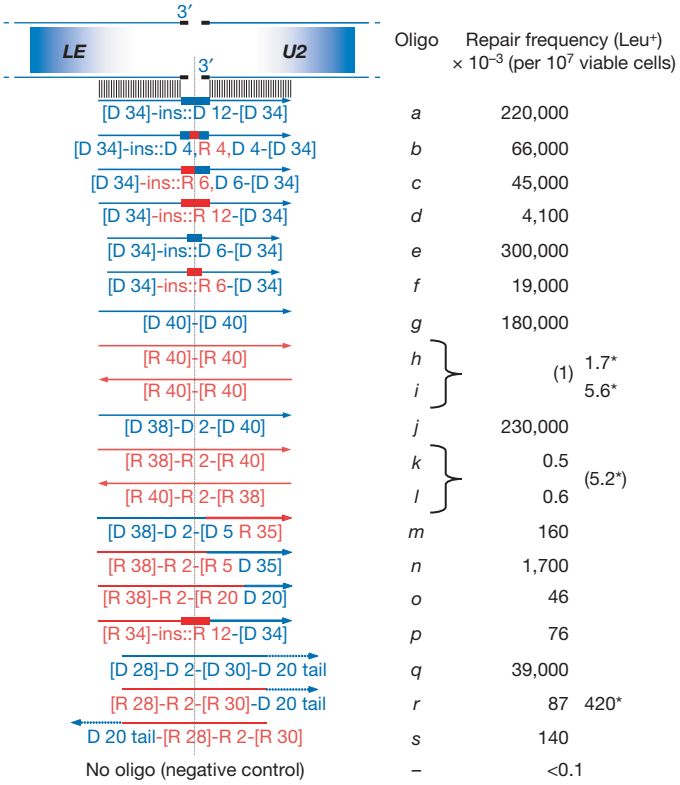


Figure 1 | Repair of a DSB by RNA-containing oligonucleotides. The diagram shows the broken *LEU2* chromosomal DNA along with the oligonucleotides containing DNA (D; blue) or RNA (R; red) sequences that were used to repair the DSB, and corresponding frequencies of *LEU2* repair. The HO cutting site (124 bp) is split in two halves shown as thicker short black lines (not to scale). Oligonucleotides are shown as lines with arrows at the 3′ end; nucleotide inserts are shown as thick lines; dotted lines indicate non-homologous tails. Potential for homologous pairing is presented as short, thin parallel vertical lines. Numbers of nucleotides homologous to the *LEU2* sequence are indicated in square brackets. Insertions are indicated by 'ins::'; a comma separates DNA from RNA bases within the insertions; a hyphen separates the different parts of the oligonucleotides. Oligonucleotide sequences are given in Supplementary Table 1. Presented are numbers of Leu$^+$ transformants per 10^7 viable cells resulting from targeting by 1 or 5* (adjacent column) nmoles of oligonucleotides *a* to *s*. Targeting frequencies with a pair of oligonucleotides (connected by braces) are shown in parentheses. Confidence intervals, as well as results of sequence verification are provided in Supplementary Table 2a.

[1]Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences (NIH, DHHS), Research Triangle Park, North Carolina 27709, USA.

replication and lead to a site-specific DSB[19]. However, the reduced transformation frequency with increased size of RNA might be due to a higher likelihood of replication arrest. If so, it is unlikely that cDNA generated by Ty reverse transcriptase of the RNA-containing oligonucleotides would be the source of DSB repair.

To exclude this possibility, that RNA-containing oligonucleotides were copied into cDNA before interacting with the DSB ends, we used our recent finding that a DSB can activate strand-biased targeting by ssDNA oligonucleotides with homology to a distant site[17]. Because many kilobases (kb) of the 5′ strand can be degraded before repair[3], there is a bias for the oligonucleotide complementary to the 3′ strand. If a cDNA intermediate were formed from ssRNA-containing oligonucleotides, the observed bias should be opposite to that with the corresponding ssDNA oligonucleotides, and there would be no bias if dsDNA were formed. As shown in Fig. 2, targeting with RNA-containing oligonucleotides was biased in favour of the oligonucleotide complementary to the 3′ end of the break, similar to DNA-only oligonucleotides. No bias was detected without DSB induction (Supplementary Table 3). Transfer of the *Bam*HI site contained in R.w and R.c (Fig. 2a) was confirmed in 28/30 transformants. We conclude that yeast cells have the ability to use RNA embedded in DNA as a template within the chromosome.

We then examined repair by RNA-only molecules that were homologous to both sides of a DSB. As shown in Fig. 1, the Leu^+ transformation frequency with five nmoles of oligonucleotides *h* or *i* reached 5×10^{-7}, and restoration of *LEU2* sequence was precise (in 26/27 clones tested). In the absence of oligonucleotides, the frequency of Leu^+ colonies was $\sim 1 \times 10^{-8}$ and in all tested isolates (16/16) the Leu^+ phenotype was due to imprecise non-homologous end joining causing small insertions or deletions (Supplementary Table 2a). Thus, ssRNA oligonucleotides are estimated to increase the precise repair of a DSB in *LEU2* by over 500-fold. Similar RNA

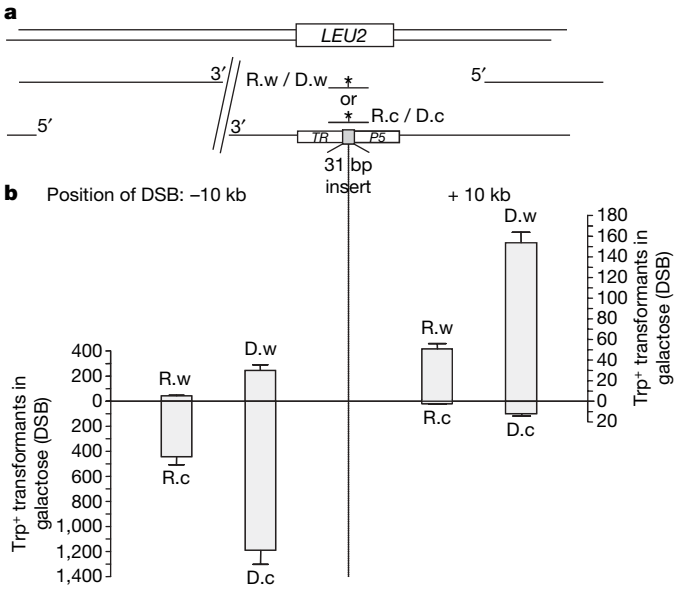


Figure 2 | Strand bias of oligonucleotide targeting to sites distant from the DSB. a, System to detect oligonucleotide strand bias in yeast diploid strains. One copy of chromosome VII with the *TRP5* locus inactivated by a 31 bp frameshift insertion plus an I-*Sce*I-induced DSB either 10 kb upstream or downstream[17]. The Trp^+ phenotype can be restored by the oligonucleotides R.w or R.c (corresponding to the 'Watson' or 'Crick' strand in the *TRP5* coding sequence), containing six central bases of RNA, or only DNA (D.w and D.c), while the intact copy of chromosome VII, in which *TRP5* is replaced by *LEU2*, provides a template for repair of the DSB. A restriction site created by the oligonucleotides is indicated by an asterisk. **b**, Number of Trp^+ transformants per 10^7 viable cells resulting from targeting 1 nmole of R.w, R.c, D.w or D.c following DSB induction. Presented are mean + s.d. from six independent experiments.

oligonucleotides containing 2-base substitutions in the centre (*k* and *l*) gave comparable Leu^+ transformation frequencies (Fig. 1). The mutations were precisely transferred (Supplementary Table 2a), indicating DNA synthesis across the RNA templates.

Several oligonucleotides were used to better understand DSB repair by RNA. The minimum size for repair by RNA-only molecules was greater than 45 bases (*k.45* and *l.45* in Supplementary Table 2a). The repair was enhanced by the presence of DNA at one end of the RNA oligonucleotides, regardless of whether the DNA was homologous (*m*, *n*, *o* and *p*) or non-homologous (*r* and *s*) to the DSB ends. Repair frequencies by oligonucleotides that required DNA synthesis through long RNA tracts were much lower than those of corresponding oligonucleotides requiring less synthesis on RNA (Fig. 1, compare *n* with *p*; also compare *b*, *c*, *d* and *f*). RNA oligonucleotides containing DNA homologous to a DSB end at their 5′ (*m*) or 3′ end (*n*) were over 100 and 1,000-fold more efficient, respectively, than the RNA-only molecules (Fig. 1). One possible explanation is that rather than simply enhancing annealing, the homologous DNA can provide a DNA duplex region close to the 3′ end of the break that could facilitate polymerase binding (compare *n* with *m* and with *o*). It is also possible that the DNA end could prevent end-degradation, especially from the 3′ end (compare *m* with *n*, and *o* with *r* and *s*) (Fig. 1). In agreement with the protection idea, the presence of a 20-base non-homologous DNA tail at the 3′ end of an oligonucleotide containing 60 nucleotides of homologous RNA (*r* and *s*), but not a 3-nucleotide tail (*r3*), increased repair by a factor of 100 when compared with the RNA-only oligonucleotides (*k* and *l*) containing an even longer RNA stretch of homology (Fig. 1 and Supplementary Table 2a). The transformation differences associated with the longer DNA tail could be due to protection from 3′→5′ exoRNases, which have a major role in RNA surveillance[20]. Deletion of the non-essential 3′→5′ exoRNase *RRP6* did not increase transformation with an RNA-only oligonucleotide (*k*) (Supplementary Table 4). However, because most 3′→5′ exoRNases are coded by essential genes in yeast, further genetic studies are required to examine the potential involvement of these enzymes. Because RNA-templated repair must involve RNA–DNA duplex intermediates, we also examined the possible role of RNases H1 (*RNH1*) and H2 (*RNH35*), which can efficiently degrade messenger RNA paired with DNA[11]. There was no increase in RNA oligonucleotide-mediated repair for the single or the double mutants (Supplementary Table 4).

Targeting of RNA oligonucleotides was stimulated by the DSB by at least 100-fold (compare *n* or *r* added to cells in galactose medium to induce the HO-endonculease versus no galactose, Supplementary Tables 2a and b) and occurred independently from the strand invasion function of Rad51 (Supplementary Table 4). The targeting was also independent of chromosomal locus because a DSB induced within *TRP5* was also precisely repaired by RNA oligonucleotides (*T2* and *T4* in Supplementary Table 2c).

Except for telomerase genes, the only genes in yeast known to code for reverse transcriptases are those contained in Ty elements. Deletion of the *SPT3* gene, which is essential for transposition and transcription of Ty1 and Ty2 elements[4,21], did not affect RNA-templated repair (Supplementary Table 5). We conclude that Ty reverse transcription has at most a minor role in DSB repair events mediated by a homologous RNA template. Deletion of telomerase genes *EST2* or *EST1* (ref. 22) also did not affect transformation by the RNA-containing oligonucleotides (Supplementary Table 5). This contrasts with the strong stimulation of Ty cDNA synthesis and transposition in an *est2*-null mutant[8].

It is also possible that DNA synthesis during RNA-templated repair is accomplished by DNA polymerases. However, deletion of individual nonessential yeast DNA polymerase genes (*POL4*, *REV1*, *REV3*, *RAD30* or *MIP1*)[23], as well as double and triple deletion mutants of translesion DNA polymerase *REV1*, *REV3* and *RAD30* genes, did not alter transformation by RNA-containing oligonucleotides (Supplementary Table 5). We conclude that the ability to

synthesize DNA on an RNA tract during repair could be a redundant function among polymerases and/or a function of one or more essential replicative DNA polymerases.

We therefore examined the ability of yeast DNA polymerases α and δ to copy templates with sequences corresponding to oligonucleotides *a* and *b* containing 0 or 4 RNA bases (Fig. 3a and Supplementary Table 1). Both polymerases copied all 4 ribonucleotides and generated full-length products (Fig. 3b; compare lanes 8 and 9, and 14 and 15). Synthesis by Pol α, but not by Pol δ (not shown), was stimulated by Mn^{2+} (Fig. 3b, lanes 2–3 versus 8–9). These polymerases also partially copied a template (IV) in which the RNA tract starts at the first single-strand template position (Fig. 3b, lanes 11 and 17). Greater extension occurred when synthesis started from a DNA tract (for example, substrate III, lanes 10 and 16, and substrate II containing a 4 ribonucleotide tract (R 4), lanes 9 and 15). Although not tested, the DNA polymerase co-factor proliferating cell nuclear antigen might enhance synthesis further, because PCNA enhances Pol δ processivity when copying DNA. On an RNA-only template (V), Pol α incorporated up to 12 bases in the presence of Mn^{2+} (lane 6), and both polymerases added a nucleotide with or without the presence of a downstream oligonucleotide (Dw), which created a gap, if Mg^{2+} was present (Fig. 3b, lanes 12 and 18, and Fig. 3c). When the efficiency of copying the 4-nucleotide template tract was determined (as described in Section 3.1 in ref. 24), Pol δ extended 80% of the products beyond R 4 and 76% of the products beyond the equivalent D 4 tract. Thus, the RNA tract was copied by Pol δ as efficiently as the corresponding DNA tract. Interestingly, once the R 4 tract was copied by Pol δ, further elongation was impeded by the presence of the RNA–DNA duplex upstream of the Pol δ active site (for example, see band highlighted with a black asterisk in Fig. 3b, lane 15). Pol α also copied the R 4 tract, but less efficiently (27% of products beyond R 4 compared with 92% of products beyond D 4). Moreover, Pol α required 20-fold more binding–synthesis-dissociation cycles (determined as described[24]) than Pol δ to completely bypass R 4. Thus, Pol δ may be the more likely candidate for reverse transcriptional repair of a DSB *in vivo*. It will be interesting to assess the effects of specific replicative DNA polymerase mutants on DSB repair by RNA-containing molecules.

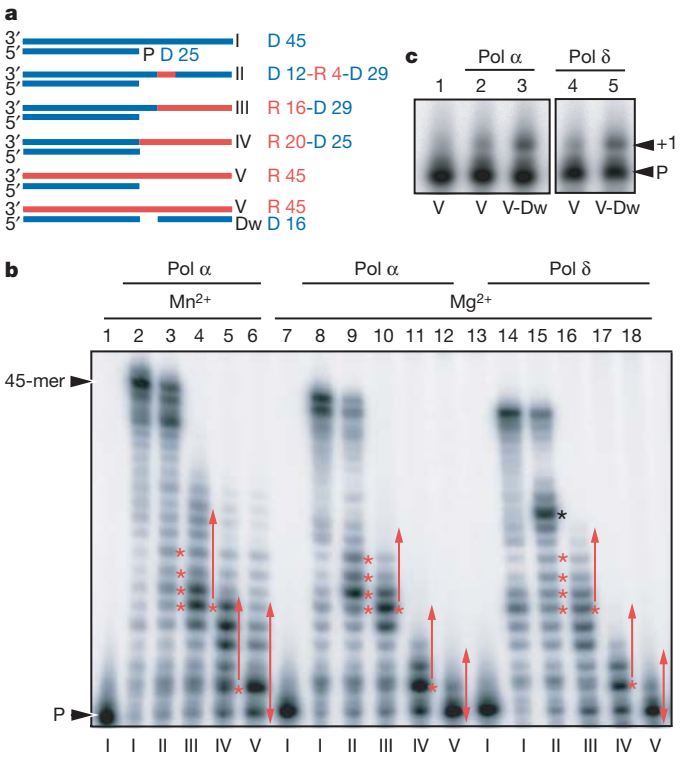


Figure 3 | Synthesis by Pol α and Pol δ across RNA templates. a, Substrates consist of DNA (D; blue) and/or RNA (R; red) and include primer P (same in all substrates). **b**, Lanes 1, 7, 13, template I, no enzyme; lanes 2–6 and 8–12, Pol α products; lanes 14–18, Pol δ products. Reactions use templates: I in lanes 2, 8, 14; II in lanes 3, 9, 15; III in lanes 4, 10 and 16; IV in lanes 5, 11, and 17; and V in lanes 6, 12, and 18. Red asterisks, ribonucleotide positions in template II and first ribonucleotide position in templates III and IV. Red arrows, RNA tracts in templates III, IV and V. The black asterisk in lane 15 marks where Pol δ was impeded when the RNA–DNA duplex is upstream of the polymerase active site. **c**, Synthesis on a gapped substrate. Lane 1, template with no enzyme V; lanes 2 and 4, products with template V and indicated enzyme; lanes 3 and 5, products with gapped substrate (V-Dw). Extension of the primer by one nucleotide is shown as +1.

The finding of *in vivo* and *in vitro* DNA synthesis on RNA-containing oligonucleotides is relevant to situations where RNA might appear within DNA *in vivo*, as shown for the mammalian mitochondrial genome[25]. The inclusion of RNA bases could occur during normal DNA metabolic reactions, as indicated by the ability of several DNA polymerases to incorporate ribonucleotides *in vitro*[25,26] and the ability of DNA ligase I to ligate RNA bases into DNA during Okazaki fragment maturation *in vitro*[27].

The RNA-templated DSB repair presented here is clearly distinct from previously described RNA/cDNA-mediated DSB repair processes[4–7,12]. We show that there is no barrier to the direct transfer of information from RNA templates to chromosomal DNA. On the basis of this newly discovered RNA capability, we suggest that endogenous RNA could have a direct role in repairing lesions during or after transcription, especially given its high local concentration. Our results set the stage for understanding how direct, homology-driven transfer of endogenous RNA information to DNA may occur.

The ability of RNA to transfer genetic information to homologous chromosomal DNA could lead to new directions in gene targeting, given that RNA can be amplified at will within cells. Moreover, RNA as a homologous template in DNA repair may contribute to both genome integrity and evolution.

METHODS

DSB induction and targeting with oligonucleotides. The RNA-containing oligonucleotides and the DNA oligonucleotides used to repair the HO or the I-*Sce*I induced DSB, or to target a sequence distant from an induced I-*Sce*I break, are described in Supplementary Table 1. DSB induction (Supplementary Methods) and targeting with oligonucleotides (1 or 5 nmoles), using a lithium acetate transformation protocol, were done as previously described[16,17,28], with the only difference being that cells to be transformed with RNA-containing oligonucleotides were washed five times with RNase-free water to dilute potential traces of RNases. All solutions and equipment used for the transformation were RNase free. Cells from each oligonucleotide transformation were plated to either selective Leu^- or Trp^- media and to synthetic complete media to determine culture viability. We excluded the possibility that DNA contamination in our RNA-containing oligonucleotides was responsible for the transfer of genetic information (Supplementary Fig. 1). Details about yeast strains are presented in the Supplementary Methods section and genetic standard methods are as described[16,28].

DNA synthesis reactions. Reactions (20 μl) with 5 nM *S. cerevisiae* Pol α (catalytic subunit), purified as described[29], contained 20 mM Tris-HCl (pH 8), 10 mM $MgCl_2$ or 0.5 mM $MnCl_2$, 2 mM DTT, 0.2 mg ml^{-1} BSA, 50 nM dNTPs and 200 nM oligonucleotide substrates prepared as described in Supplementary Table 1. Reactions with the 3-subunit *S. cerevisiae* Pol δ (gift from P. M. J. Burgers) were the same except for the use of 40 mM Tris (pH 8), 5 mM $MgCl_2$, 0.1 mg ml^{-1} BSA and 75 mM NaCl. After a 6 min incubation at 30 °C, reaction mixtures were quenched by adding 20 μl of 99% formamide, 5 mM EDTA, 0.1% xylene cyanole, 0.1% bromophenol blue, resolved by electrophoresis in a 12% denaturing polyacrylamide gel and visualized using a Molecular Dynamics PhosphorImager.

Received 27 September 2006; accepted 28 February 2007.
Published online 11 April 2007.

1. Baltimore, D. Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. *Cell* **40**, 481–482 (1985).
2. Autexier, C. & Lue, N. F. The structure and function of telomerase reverse transcriptase. *Annu. Rev. Biochem.* **75**, 493–517 (2006).

3. Paques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63,** 349–404 (1999).
4. Teng, S. C., Kim, B. & Gabriel, A. Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature* **383,** 641–644 (1996).
5. Moore, J. K. & Haber, J. E. Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. *Nature* **383,** 644–646 (1996).
6. Derr, L. K. & Strathern, J. N. A role for reverse transcripts in gene conversion. *Nature* **361,** 170–173 (1993).
7. Nevo-Caspi, Y. & Kupiec, M. cDNA-mediated Ty recombination can take place in the absence of plus-strand cDNA synthesis, but not in the absence of the integrase protein. *Curr. Genet.* **32,** 32–40 (1997).
8. Lesage, P. & Todeschini, A. L. Happy together: the life and times of Ty retrotransposons and their hosts. *Cytogenet. Genome Res.* **110,** 70–90 (2005).
9. Kasahara, M., Clikeman, J. A., Bates, D. B. & Kogoma, T. RecA protein-dependent R-loop formation *in vitro*. *Genes Dev.* **14,** 360–365 (2000).
10. Zaitsev, E. N. & Kowalczykowski, S. C. A novel pairing process promoted by *Escherichia coli* RecA protein: inverse DNA and RNA strand exchange. *Genes Dev.* **14,** 740–749 (2000).
11. Huertas, P. & Aguilera, A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* **12,** 711–721 (2003).
12. Morrish, T. A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature Genet.* **31,** 159–165 (2002).
13. Ricchetti, M. & Buc, H. E. *coli* DNA polymerase I as a reverse transcriptase. *EMBO J.* **12,** 387–396 (1993).
14. Murakami, E. *et al.* Characterization of novel reverse transcriptase and other RNA-associated catalytic activities by human DNA polymerase γ: importance in mitochondrial DNA replication. *J. Biol. Chem.* **278,** 36403–36409 (2003).
15. Franklin, A., Milburn, P. J., Blanden, R. V. & Steele, E. J. Human DNA polymerase-η, an A-T mutator in somatic hypermutation of rearranged immunoglobulin genes, is a reverse transcriptase. *Immunol. Cell Biol.* **82,** 219–225 (2004).
16. Storici, F., Durham, C. L., Gordenin, D. A. & Resnick, M. A. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc. Natl Acad. Sci. USA* **100,** 14994–14999 (2003).
17. Storici, F., Snipe, R. J., Chan, G. K., Gordenin, D. A. & Resnick, M. A. Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing. *Mol. Cell. Biol.* **26,** 7645–7657 (2006).
18. Paques, F., Leung, W. Y. & Haber, J. E. Expansions and contractions in a tandem repeat induced by double-strand break repair. *Mol. Cell. Biol.* **18,** 2045–2054 (1998).
19. Vengrova, S. & Dalgaard, J. Z. The wild-type *Schizosaccharomyces pombe mat1* imprint consists of two ribonucleotides. *EMBO Rep.* **7,** 59–65 (2006).
20. Houseley, J., LaCava, J. & Tollervey, D. RNA-quality control by the exosome. *Nature Rev. Mol. Cell Biol.* **7,** 529–539 (2006).
21. Boeke, J. D., Styles, C. A. & Fink, G. R. *Saccharomyces cerevisiae SPT3* gene is required for transposition and transpositional recombination of chromosomal Ty elements. *Mol. Cell. Biol.* **6,** 3575–3581 (1986).
22. Smogorzewska, A. & de Lange, T. Regulation of telomerase by telomeric proteins. *Annu. Rev. Biochem.* **73,** 177–208 (2004).
23. Shcherbakova, P. V. & Kunkel, T. A. in *DNA Replication in Eukaryotic Cells and its Relevance to Human Disease* (ed. DePamphilis M.L.) 391–410 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2006).
24. McCulloch, S. D. & Kunkel, T. A. Multiple solutions to inefficient lesion bypass by T7 DNA polymerase. *DNA Repair* **5,** 1373–1383 (2006).
25. Yang, M. Y. *et al.* Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. *Cell* **111,** 495–505 (2002).
26. Nick McElhinny, S. A. & Ramsden, D. A. Polymerase mu is a DNA-directed DNA/RNA polymerase. *Mol. Cell. Biol.* **23,** 2309–2315 (2003).
27. Rumbaugh, J. A., Murante, R. S., Shi, S. & Bambara, R. A. Creation and removal of embedded ribonucleotides in chromosomal DNA during mammalian Okazaki fragment processing. *J. Biol. Chem.* **272,** 22591–22599 (1997).
28. Storici, F. & Resnick, M. A. The *delitto perfetto* approach to *in vivo* site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol.* **409,** 329–345 (2006).
29. Niimi, A. *et al.* Palm mutants in DNA polymerases α and η alter DNA replication fidelity and translesion activity. *Mol. Cell. Biol.* **24,** 2734–2746 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. E. Haber for yeast strain YFP17 and P. M. J. Burgers for yeast DNA polymerase δ. We thank C. Halweg and W. C. Copeland for suggestions; K. L. Adelman, J. W. Drake and A. Sugino for critical reading of the manuscript; and J. R. Snipe and G. K. Chan for technical assistance. Research support was from National Institute of Environmental Health Sciences (NIH) intramural research funds.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.R. (resnick@niehs.nih.gov).

Vol 447|17 May 2007|doi:10.1038/nature05702

LETTERS

Structural basis for cofactor-independent dioxygenation in vancomycin biosynthesis

Paul F. Widboom[1]*, Elisha N. Fielding[1]*, Ye Liu[1] & Steven D. Bruner[1]

Enzyme-catalysed oxidations are some of the most common transformations in primary and secondary metabolism. The vancomycin biosynthetic enzyme DpgC belongs to a small class of oxygenation enzymes that are not dependent on an accessory cofactor or metal ion[1]. The detailed mechanism of cofactor-independent oxygenases has not been established. Here we report the first structure of an enzyme of this oxygenase class in complex with a bound substrate mimic. The use of a designed, synthetic substrate analogue allows unique insights into the chemistry of oxygen activation. The structure confirms the absence of cofactors, and electron density consistent with molecular oxygen is present adjacent to the site of oxidation on the substrate. Molecular oxygen is bound in a small hydrophobic pocket and the substrate provides the reducing power to activate oxygen for downstream chemical steps. Our results resolve the unique and complex chemistry of DpgC, a key enzyme in the biosynthetic pathway of an important class of antibiotics[2]. Furthermore, mechanistic parallels exist between DpgC and cofactor-dependent flavoenzymes[3], providing information regarding the general mechanism of enzymatic oxygen activation.

Aerobic organisms exploit the reactivity of molecular oxygen for critical processes such as energy production and the synthesis of cellular components. The direct reaction of molecular oxygen with organic molecules is formally a disallowed, spin-forbidden transformation. The vast majority of characterized oxygenases use bound transition metals or flavin cofactors to activate triplet molecular oxygen to carry out diverse oxidation chemistry[4]. The vancomycin biosynthetic enzyme DpgC is a unique catalyst that performs a dioxygenation reaction independent of metals or cofactors[1].

A wide variety of therapeutically important natural products are synthesized by non-ribosomal peptide synthetases[5]. Diverse non-proteinogenic amino acid building blocks allow the construction of complex peptide architectures that impart unique biological activity. The enzyme DpgC catalyses a key step in the biosynthesis of 3,5-dihydroxyphenylglycine (DPG), a non-proteinogenic amino acid found in the vancomycin and teicoplanin family of antibiotics[6] (Fig. 1). DPG is an essential building block in this family, forming biaryl crosslinks that give a rigid structure necessary for biological activity. The biosynthetic pathway for DPG is composed of five enzymes in the producer actinomycetes bacteria[7]. The enzymes DpgA, DpgB and DpgD couple four malonyl-CoAs to form 3,5-dihydroxyphenylacetate-CoA (DPA-CoA). The dioxygenase DpgC next oxidizes and cleaves DPA-CoA to DPGX (3,5-dihydroxyphenylglyoxylate) and coenzyme A (CoA). Biochemical studies have clearly established that DpgC is a dioxygenase, incorporating two oxygen atoms from the same molecule of molecular oxygen into the substrate[1]. DpgC performs both the observed four-electron oxidation and the hydrolysis of the thioester bond, without a metal or cofactor. This combination of transformations has no precedent in other enzymes.

The carboxy terminus of DpgC shows homology to the crotonase family of enoyl-CoA isomerase/dehydratases[8]. Members of the crotonase family perform diverse chemical transformations on acyl-coenzyme A substrates[9]. However, there are no examples of oxidation/reduction chemistry from this enzyme class. The amino-terminal region of DpgC, comprising approximately one-third of the enzyme, is neither homologous to crotonase enzymes nor to any deposited protein sequence. In addition, assigning catalytic residues in DpgC on the basis of sequence homology to the crotonase superfamily is not obvious. The deposited DpgC protein sequences from various glycopeptide gene clusters are all highly homologous to one another (~70% identical and 80% similar at the amino acid level)[6].

We crystallized apo-DpgC in the *R*3 space group. The asymmetric unit contains four monomers; two monomer pairs each form a hexamer through crystallographic symmetry operations. As confirmed experimentally, DpgC exists in solution as a hexamer (refer to Supplementary Fig. 3). The structure was solved through single-wavelength anomalous dispersion (SAD) phasing of a selenomethionine derivative to 2.75 Å resolution. The structure of the apo-enzyme contains numerous disordered regions, especially in proximity to the predicted active site. To resolve the problems of disorder in the crystal and gain mechanistic insight into the chemistry of DpgC, we designed and synthesized a substrate analogue for use in co-crystallization studies.

The coenzyme A analogue, DPA-NH-CoA, (Fig. 1, **4**) was designed to be a stable, isosteric mimic of the substrate DPA-CoA, containing an amide in place of the natural thioester. Formation of an

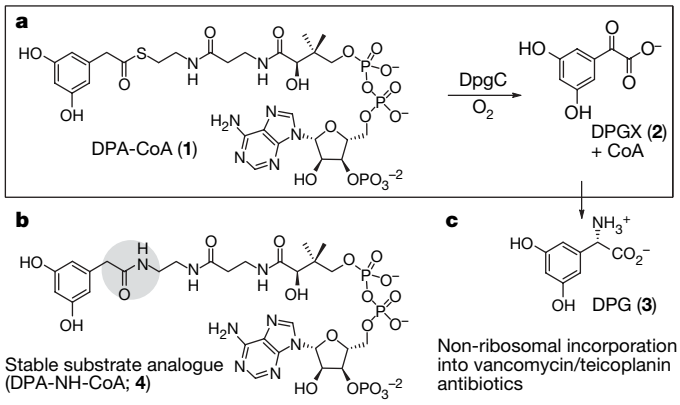


Figure 1 | The dioxygenase DpgC. a, The conversion of DPA-CoA (**1**) to DPGX (**2**) and coenzyme A (CoA) catalysed by DpgC. **b**, Structure of the amide analogue of DPA-CoA (DPA-NH-CoA, **4**), designed as a stable substrate mimic and inhibitor of DpgC. **c**, Product of the biosynthetic pathway, the non-proteinogenic amino acid 3,5-dihydroxyphenylglycine (DPG, **3**).

[1]Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, Massachusetts 02467, USA.
*These authors contributed equally to this work.

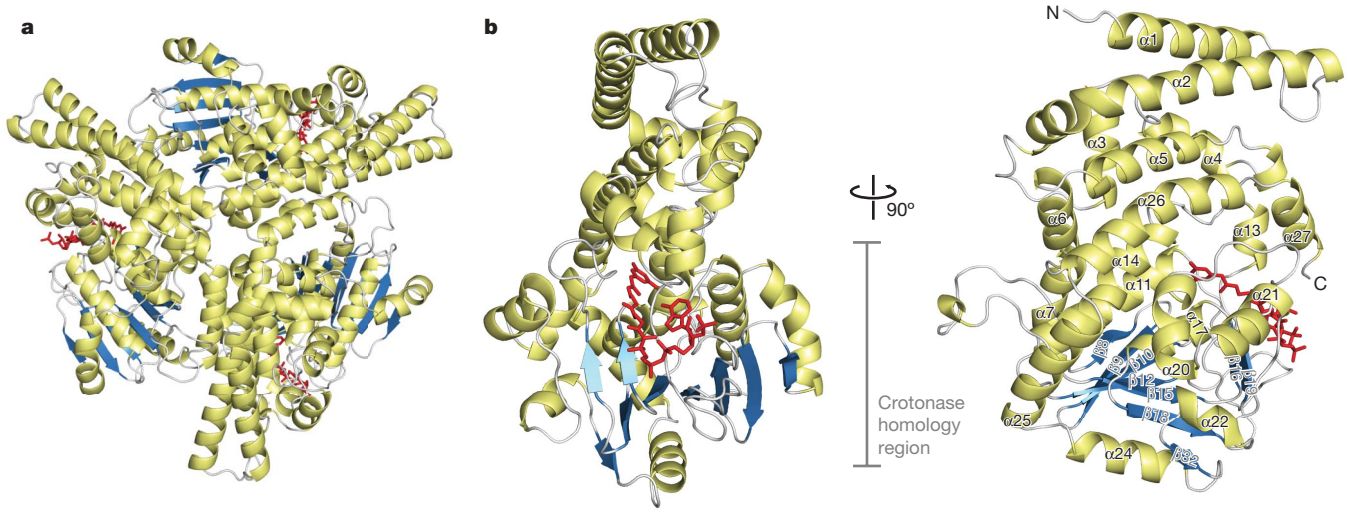


Figure 2 | The X-ray structure of DpgC. a, Ribbon diagram of DpgC looking down the three-fold axis of the hexamer; α-helices are coloured yellow and β-sheets blue. The bound inhibitor DPA-NH-CoA is shown in red. **b**, Two views of the DpgC monomer with the C-terminal crotonase homology region indicated.

acyl-thioester enol/enolate is predicted to be the first step in the catalytic cycle of DpgC[8]. We reasoned that deprotonation of the amide α-hydrogens in DPA-NH-CoA would be blocked, preventing enolization and subsequent downstream chemistry[10]. DPA-NH-CoA was synthesized and its K_i determined to be 2.7 μM (see Supplementary Information), comparable to the Michaelis constant (K_m) of 3.9 μM for the natural substrate and that measured for DpgC from the chloroeremomycin gene cluster (6 μM)[7]. The similarity between the K_m of the natural substrate and the K_i of our inhibitor is consistent with **4** acting as a substrate mimic. In addition, under all conditions tested, DPA-NH-CoA was recovered unchanged during enzyme turnover conditions.

The complex between DpgC and DPA-NH-CoA crystallized in the space group $P2_12_12$ with two trimers per asymmetric unit, each forming a dimer of trimers in the unit cell (Fig. 2). The active site is well ordered, and the substrate analogue, DPA-NH-CoA, is clearly evident in the electron density maps. The coenzyme A portion of DPA-NH-CoA binds in the active site with an extended conformation and the adenosine moiety bent back around the molecule (Fig. 2b). The 3,5-dihydroxyphenyl ring is bound in a well-defined pocket with a hydrogen bond between the 5-hydroxyl and an ordered water molecule (W 141; red sphere, Fig. 3a). The 3-hydroxyl forms a hydrogen bond with Glu 189 and interacts with the face of Arg 254. This is an unusual interaction because the phenolic oxygen is nearest (3.3 Å) to the carbon of the Arg 254 guanidinium group (Fig. 3a). The amide of the substrate analogue DPA-NH-CoA is properly oriented in an 'oxyanion hole', similar in architecture to that previously described for other members of the crotonase superfamily[11]. The backbone amide hydrogens of Ile 235 and Gly 296 are positioned with a favourable geometry to stabilize the predicted intermediate thioester enolate (Fig. 3b). An ordered water (W 258, Fig. 3b) is properly positioned (3.2 Å from the benzylic carbon) to assist in the stereospecific removal of the proR (Fig. 4) hydrogen of the substrate. The water is coordinated to a hydrogen bonding network, leading to bulk solvent anchored by hydrogen bonds with Gln 299. Enzyme stabilization of the enolate form of the acyl-CoA substrate favours deprotonation, shifting the equilibrium to the enolate anion. A specific enzyme-associated 'general base' is probably not necessary on the basis of the increased acidity of the α-hydrogens in the enzyme-bound substrate.

Deprotonation of the proR hydrogen of the substrate suggests that molecular oxygen will react from the opposite face of the substrate. At this predicted position, we observed unassigned electron density consistent in size and shape with diatomic molecular oxygen. The observed electron density is positioned above the proS hydrogen of the substrate. Each of the three DpgC monomers in the asymmetric unit of the crystal contains density in the same location and orientation relative to the substrate mimic. Dioxygen models especially well into the density map of one of the three active sites (Fig. 3c). Ordered molecular oxygen is rarely observed in protein crystal structures, especially when not complexed to a transition metal[12,13]. The observed electron density, which we postulate to be bound oxygen, is in a well-defined, solvent accessible hydrophobic pocket lined with residues Leu 237, Ile 324, Val 425 and Val 429. There is a single hydrogen bond donor (backbone NH of Ile 324) and no hydrophilic residues in the pocket. The size of the orphan electron density and the hydrophobic nature of the binding pocket suggest it is not water or a

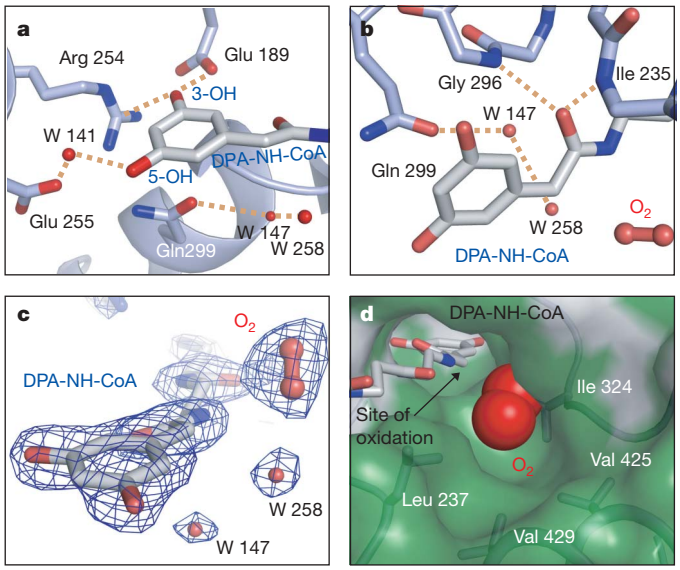


Figure 3 | Substrate recognition and active site of DpgC. a, Recognition of the 3,5-dihydroxyphenyl ring of the substrate DPA-CoA. **b**, The active site of DpgC with the bound substrate analogue DPA-NH-CoA. Residues Gly 296 and Ile 235 comprise the 'oxyanion hole'. **c**, Composite omit electron density map (contoured at 1.7σ) showing the area around the substrate analogue, the catalytically relevant water (W 258) and molecular oxygen. **d**, Surface representation of the hydrophobic oxygen-binding pocket. Hydrophobic surfaces are coloured green and molecular oxygen is shown in van der Waals representation (red).

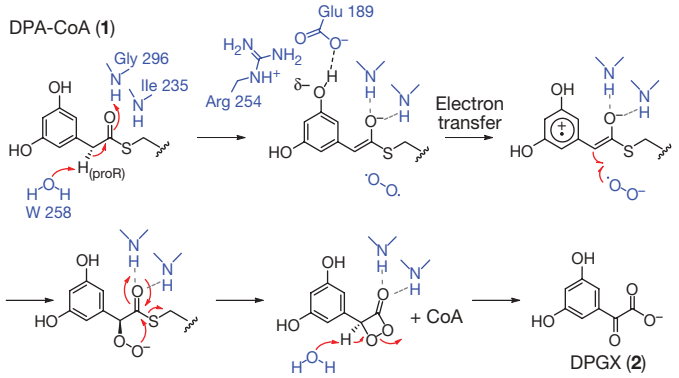


Figure 4 | Proposed catalytic mechanism. The structures of the substrate DPA-CoA and amino acids are abbreviated to show only sections relevant to the chemistry.

cationic metal that is bound. In addition, the crystallization media contained no compounds that could fit into the binding pocket. Thus, oxygen is the logical candidate for the unassigned electron density that we observe. Molecular oxygen models within the van der Waals contact range of the hydrophobic residues and 3.3 Å from the backbone amide hydrogen bond donor.

Unequivocal assignment of the electron density as molecular oxygen is not possible based on analysis of the diffraction data alone. The noble gas xenon has been exploited as an isoelectronic/isosteric mimic of O_2 (ref. 14). Soaking of our co-complex crystals under high pressures of xenon gas failed to displace the observed bound oxygen, suggesting the oxygen is bound with unique specificity. To understand the nature of the hydrophobic pocket, we used site-directed mutagenesis to alter subtly the hydrophobicity of the putative oxygen-binding site. We mutated the four hydrophobic residues (Fig. 3d) to threonine, thus maintaining in part the hydrophobic character of the side chains while adding an electronegative hydroxyl group. These minor changes, far from the substrate binding site, should disrupt oxygen binding and reduce enzyme activity if our model is accurate. As illustrated in Table 1, all four mutations affect the dioxygenase activity of DpgC as determined by the k_{cat}/K_m values of the substrate. In addition, the K_m values of molecular oxygen, as measured with an oxygen electrode, are consistent with our mechanistic hypothesis. Isoleucine 324 seems especially critical for enzyme activity because mutation to threonine abolishes activity.

In addition to fitting in the observed electron density, an ordered dioxygen molecule at the observed position is consistent with the unusual chemistry of DpgC. Our proposed mechanism is illustrated in Fig. 4. The first step of the reaction is formation of a thioester enolate. The backbone amides of residues Gly 296 and Ile 235 are properly positioned to activate the substrate and stabilize the thio-enolate oxyanion. Deprotonation, facilitated by an ordered water molecule (W 258), is predicted to be stereospecific, giving a highly conjugated anion intermediate. Molecular oxygen, bound in a hydrophobic pocket and oriented by a hydrogen bond from the backbone NH of Ile 324, is properly positioned to react with the substrate. Reaction with molecular oxygen is a two-step process.

Table 1 | Effect of mutations in the hydrophobic oxygen-binding pocket on the catalytic activity of DpgC

	K_m (DPA-CoA)	k_{cat} (DPA-CoA)	k_{cat}/K_m*	K_m (dioxygen)
Wild type	3.9	0.172	1.0	1.7
Ile324Thr	>1,000	ND	ND	ND
Leu237Thr	58.0	0.050	0.019	2.1
Val429Thr	14.7	0.019	0.029	>2.6
Val425Thr	13.8	0.037	0.060	2.6

The units for $K_{m\ (DPA\text{-}CoA)}$ are μM, k_{cat} are in s^{-1} and $K_{m(dioxygen)}$ are mM. ND, not determined.
*Apparent k_{cat}/K_m relative to the wild-type enzyme.

The first step is the transfer of an electron from the substrate to triplet oxygen, forming a conjugated radical cation/superoxide pair. The observed interactions of Glu 189 and Arg 254 with the substrate are consistent with stabilization of a partial (or perhaps full) negative charge on the 3-hydroxyl, adding additional electron density to the substrate intermediate and facilitating the electron transfer step. The pair can then collapse in a spin-allowed, bond-forming process to give the peroxide intermediate. With the thioester still positioned in the activating oxyanion hole, collapse to form a 1,2-dioxetanone as well as elimination of coenzyme A is facile and consistent with model chemical systems[15]. The α-hydrogen of the 1,2-dioxetanone is in a similar position to the proR hydrogen in substrate DPA-CoA, allowing deprotonation analogous to the first step, and generating the observed product, DPGX. The enzyme must carefully direct the chemistry of the 1,2-dioxetanone intermediate to prevent decomposition to carbon dioxide as observed in firefly luciferase[16].

Biochemical precedent for the proposed mechanism is found in a small subset of enzyme-catalysed reactions that employ either an aromatic substrate or cofactor[17]. X-ray structures of the metal-free oxygenases, urate oxidase and aromatic monooxygenase involved in polyketide biosynthesis have been determined, but the mechanism of oxygen binding and activation is not clear[18–20]. There is biochemical evidence for the formation of an organic radical anion and superoxide by the α,β-hydrolase-fold family members involved in the anthranilate pathway of quinaldine degradation[21]. In addition, the proposed mechanism of DpgC shows similarities to the reaction of reduced flavin cofactors with molecular oxygen, which is also believed to proceed through a two-step single-electron reaction[22]. The proposed function of a histidine residue for the stabilization of an anionic flavin cofactor is analogous to the role suggested for Arg 254 in DpgC[23]. However, there are no structures of flavoenzymes with molecular oxygen bound in the reactive conformation. The presented structure of DpgC provides detailed information regarding the specific interaction with an oxygenase and a bound, isolated oxygen molecule. In addition, the structure reveals novel insights that can be applied to the general mechanism of enzymatic oxygen activation.

METHODS

Protein cloning/purification, site-directed mutagenesis, biochemical assays and synthetic methodology are described in Supplementary Methods.

Protein crystallization. Apo-DpgC (48 kDa, 439 amino acids) from the *Streptomyces toyocaensis* A47934 gene cluster[24] was crystallized by the hanging-drop method at 20 °C. DpgC (1.5 μl of 10 mg ml^{-1} in 20 mM Tris HCl and 50 mM NaCl, pH 7.5) was mixed with 1.5 μl of reservoir solution (95 mM Na HEPES, 0.94 M NaCl and 1.3 M $(NH_4)_2SO_4$, pH 7.5). Crystals appeared within two days and were allowed to continue growth for several additional days. Crystals were transferred to a cryoprotectant solution of the reservoir solution with 20% glycerol and soaked for 30 min before being flash frozen in liquid nitrogen. Co-crystals of DpgC with substrate analogue DPA-NH-CoA formed under distinct conditions. DPA-NH-CoA (2 mM) was incubated with DpgC (12 mg ml^{-1} in 20 mM Tris HCl and 50 mM NaCl, pH 7.5) at 20 °C for 2 h. The DpgC/inhibitor complex (1.5 μl) was mixed with 1.5 μl of reservoir solution: 100 mM sodium citrate, 165 mM ammonium acetate, and 24% (w/v) PEG 4000, pH 5.6. Crystals appeared after two days at 20 °C, and were transferred to a cryoprotectant solution (reservoir solution with 20% glycerol) and soaked for 30 min before being flash frozen in liquid nitrogen.

Structural determination. Selenomethionine-derivatized DpgC failed to yield diffraction-quality crystals under conditions similar to those used for the native protein. A mutant enzyme was constructed using site-directed mutagenesis (Stratagene Quikchange Kit) to alter three non-conserved methionines. The triple mutant (Met180Leu, Met252Leu, Met409Leu) selenomethionine protein crystallized under the same conditions and space group as the apo-enzyme. SAD X-ray diffraction data were collected on the X4A beamline at the National Synchrotron Light Source at Brookhaven National Labs. All diffraction data were indexed, integrated and scaled using the HKL2000 program[25]. Heavy atom phasing was carried out with the PHENIX software package[26] using 19 of the possible 28 selenium atoms with an overall figure of merit of 0.26 and anomalous signal of 6.33%. Initial electron density maps were calculated with the SAD phases from PHENIX at 2.75 Å. The crystals belong to the *R*3 space group with four

monomers per asymmetric unit. The quality of the initial maps was improved by solvent flattening and non-crystallographic averaging. Model building was done using the program COOT[27] and initial structure refinement performed with the program CNS[28]. The apo-structure contained numerous regions where electron density was not evident and building a complete model was not feasible. X-ray diffraction data of the DpgC/inhibitor co-complex crystals were collected on the X26C beamline at Brookhaven National Labs. Phases for the co-complex structure were calculated with the program AMORE[29] using the incomplete native structure as a molecular replacement search model. Initial structural refinement was performed with NCS restraints, and after several rounds the restraints were removed from the calculations. Sigma-weighted simulated annealing composite omit maps (CNS) were used to judge and verify structures throughout refinement (Supplementary Fig. 1). On the basis of examination of the asymmetric unit, two trimers are present (Matthews coefficient of 3.21 and 62% solvent content calculated for 6 monomers). However, only one of the two trimers was built and refined because electron density corresponding to the second trimer was weak and not interpretable. All crystals screened had this property. As a result of building half the predicted model, the *R*-factors are characteristically high at 2.45 Å resolution ($R_{working} = 32\%$ and $R_{free} = 36\%$). The final model contains residues 11–432 in monomers A and B, and 12–432 in monomer C.

Received 17 January; accepted 23 February 2007.

1. Tseng, C. C., Vaillancourt, F. H., Bruner, S. D. & Walsh, C. T. DpgC is a metal- and cofactor-free 3,5-dihydroxyphenylacetyl-CoA 1,2-dioxygenase in the vancomycin biosynthetic pathway. *Chem. Biol.* **11,** 1195–1203 (2004).
2. Walsh, C. T. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303,** 1805–1810 (2004).
3. Eswaramoorthy, S., Bonanno, J. B., Burley, S. K. & Swaminathan, S. Mechanism of action of a flavin-containing monooxygenase. *Proc. Natl Acad. Sci. USA* **103,** 9832–9837 (2006).
4. Silverman, R. B. (ed.) *The Organic Chemistry of Enzyme-Catalyzed Reactions*. Revised edn. (Academic Press, San Diego, 2002).
5. Cane, D. E., Walsh, C. T. & Khosla, C. Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282,** 63–68 (1998).
6. Hubbard, B. K. & Walsh, C. T. Vancomycin assembly: nature's way. *Angew. Chem. Int. Edn Engl.* **42,** 730–765 (2003).
7. Chen, H., Tseng, C. C., Hubbard, B. K. & Walsh, C. T. Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (*S*)-3,5-dihydroxyphenylglycine. *Proc. Natl Acad. Sci. USA* **98,** 14901–14906 (2001).
8. Agnihotri, G. & Liu, H. W. Enoyl-CoA hydratase. reaction, mechanism, and inhibition. *Bioorg. Med. Chem.* **11,** 9–20 (2003).
9. Holden, H. M., Benning, M. M., Haller, T. & Gerlt, J. A. The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme A thioesters. *Acc. Chem. Res.* **34,** 145–157 (2001).
10. Beaman, T. W. *et al.* Acyl group specificity at the active site of tetrahydridipicolinate *N*-succinyltransferase. *Protein Sci.* **11,** 974–979 (2002).
11. Xiang, H., Luo, L., Taylor, K. L. & Dunaway-Mariano, D. Interchange of catalytic activity within the 2-enoyl-coenzyme A hydratase/isomerase superfamily based on a common active site template. *Biochemistry* **38,** 7638–7652 (1999).
12. Berglund, G. I. *et al.* The catalytic pathway of horseradish peroxidase at high resolution. *Nature* **417,** 463–468 (2002).
13. Karlsson, A. *et al.* Crystal structure of naphthalene dioxygenase: side-on binding of dioxygen to iron. *Science* **299,** 1039–1042 (2003).
14. Wentworth, P. Jr *et al.* Antibody catalysis of the oxidation of water. *Science* **293,** 1806–1811 (2001).
15. Adam, W., Alzerreca, A., Liu, J. C. & Yany, F. Cyclic peroxide 52. alpha-peroxylactones via dehydrative cyclization of alpha-hydroperoxy acids. *J. Am. Chem. Soc.* **99,** 5768–5773 (1977).
16. Koo, J. A., Schmidt, S. P. & Schuster, G. B. Bioluminescence of the firefly: key steps in the formation of the electronically excited state for model systems. *Proc. Natl Acad. Sci. USA* **75,** 30–33 (1978).
17. Fetzner, S. Oxygenases without requirement for cofactors or metal ions. *Appl. Microbiol. Biotechnol.* **60,** 243–257 (2002).
18. Sciara, G. *et al.* The structure of ActVA-Orf6, a novel type of monooxygenase involved in actinorhodin biosynthesis. *EMBO J.* **22,** 205–215 (2003).
19. Colloc'h, N. *et al.* Crystal structure of the protein drug urate oxidase-inhibitor complex at 2.05 A resolution. *Nature Struct. Biol.* **4,** 947–952 (1997).
20. Beinker, P. *et al.* Crystal structures of SnoaL2 and AclR: two putative hydroxylases in the biosynthesis of aromatic polyketide antibiotics. *J. Mol. Biol.* **359,** 728–740 (2006).
21. Frerichs-Deeken, U. *et al.* Dioxygenases without requirement for cofactors and their chemical model reaction: compulsory order ternary complex mechanism of 1*H*-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase involving general base catalysis by histidine 251 and single-electron oxidation of the substrate dianion. *Biochemistry* **43,** 14485–14499 (2004).
22. Massey, V. Activation of molecular oxygen by flavins and flavoproteins. *J. Biol. Chem.* **269,** 22459–22462 (1994).
23. Roth, J. P. & Klinman, J. P. Catalysis of electron transfer during activation of O_2 by the flavoprotein glucose oxidase. *Proc. Natl Acad. Sci. USA* **100,** 62–67 (2003).
24. Pootoolal, J. *et al.* Assembling the glycopeptide antibiotic scaffold: The biosynthesis of A47934 from *Streptomyces toyocaensis* NRRL15009. *Proc. Natl Acad. Sci. USA* **99,** 8962–8967 (2002).
25. Otwinowski, Z. & Minor, W. in *Methods in Enzymology: Macromolecular Crystallography* Vol. 276 (eds Sweet, R. M. & Carter, C. W.) Part A, 307–326 (Academic Press, New York, 1997).
26. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
27. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**(Pt 12 Pt 1), 2126–2132 (2004).
28. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
29. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to A. Heroux and the staff at the Brookhaven National Synchrotron Light Source PXRR for assistance with X-ray data collection and processing. We are grateful to C. C. Tseng, R. S. Klausen and C. T. Walsh for discussions and experimental assistance. We thank P. D. Fortin and A. R. Howard-Jones for assistance in oxygen-binding measurements, Z. Dauter for advice on X-ray data analysis, and members of the Bruner research group and Boston College Chemistry Department for helpful discussions and comments on the manuscript. This work was supported with funds from Boston College and the Damon Runyon Cancer Research Foundation.

Author Contributions P.F.W. crystallized and solved the structure of apo-DpgC. E.N.F. crystallized the DpgC/inhibitor co-complex and with P.F.W. solved the structure. E.N.F constructed and with P.F.W. performed biochemical assays on mutant DpgC enzymes. Y.L. synthesized and biochemically evaluated the inhibitor, DPA-NH-CoA. S.D.B. wrote the manuscript and all authors discussed the results and commented on the manuscript.

Author Information The atomic coordinates for the DpgC/inhibitor complex structure have been deposited in the Protein Data Bank with accession code 2NP9. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.D.B. (bruner@bc.edu).

ADDENDUM

doi:10.1038/nature05789

Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase

Helen E. Bryant, Niklas Schultz, Huw D. Thomas, Kayan M. Parker, Dan Flower, Elena Lopez, Suzanne Kyle, Mark Meuth, Nicola J. Curtin & Thomas Helleday

Nature 434, 913–917 (2005)

Two patent applications (patent numbers US 2005/0143370 and WO 2005/012524, filed in July 2004) with relevance to this work have been filed by Cancer Research Technology Limited and The University of Sheffield and we want to declare our relationship with these patents. N.J.C. and T.H., and T.H. alone, are named inventors on these patents, respectively, but are not owners of the patents nor do they have shares in the companies or institutions to which the patents belong. However, because the value of these patents has increased, owing to further results obtained in other laboratories since publication, N.J.C. and T.H. now wish to declare possible competing financial interests.

CORRIGENDUM

doi:10.1038/nature05882

Sheep don't forget a face

Keith M. Kendrick, Ana P. da Costa, Andrea E. Leigh, Michael R. Hinton & Jon W. Peirce

Nature 414, 165–166 (2001)

In Fig. 1b of this Brief Communication, the legend and text did not make it clear that two different groups of ten sheep were used in the study to give overall $n = 20$. A reanalysis of the data using a post-hoc Tukey test (rather than a paired *t*-test, as originally stated) revealed some small errors that altered the significance values slightly; however, there is no overall change in the results. The maximum retest interval was 801 rather than 800 days, and 100–500 trials were conducted for 1–6 weeks (not 400–500 trials for 4–6 weeks, as published). A revised version of Fig. 1b showing the corrected statistical changes and an expanded legend incorporating further methodological detail are available as Supplementary Information to this Corrigendum.

Supplementary Information is linked to the online version of this Corrigendum at www.nature.com/nature.

ADDENDUM

doi:10.1038/nature05789

Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase

Helen E. Bryant, Niklas Schultz, Huw D. Thomas, Kayan M. Parker, Dan Flower, Elena Lopez, Suzanne Kyle, Mark Meuth, Nicola J. Curtin & Thomas Helleday

Nature 434, 913–917 (2005)

Two patent applications (patent numbers US 2005/0143370 and WO 2005/012524, filed in July 2004) with relevance to this work have been filed by Cancer Research Technology Limited and The University of Sheffield and we want to declare our relationship with these patents. N.J.C. and T.H., and T.H. alone, are named inventors on these patents, respectively, but are not owners of the patents nor do they have shares in the companies or institutions to which the patents belong. However, because the value of these patents has increased, owing to further results obtained in other laboratories since publication, N.J.C. and T.H. now wish to declare possible competing financial interests.

CORRIGENDUM

doi:10.1038/nature05882

Sheep don't forget a face

Keith M. Kendrick, Ana P. da Costa, Andrea E. Leigh, Michael R. Hinton & Jon W. Peirce

Nature 414, 165–166 (2001)

In Fig. 1b of this Brief Communication, the legend and text did not make it clear that two different groups of ten sheep were used in the study to give overall $n = 20$. A reanalysis of the data using a post-hoc Tukey test (rather than a paired *t*-test, as originally stated) revealed some small errors that altered the significance values slightly; however, there is no overall change in the results. The maximum retest interval was 801 rather than 800 days, and 100–500 trials were conducted for 1–6 weeks (not 400–500 trials for 4–6 weeks, as published). A revised version of Fig. 1b showing the corrected statistical changes and an expanded legend incorporating further methodological detail are available as Supplementary Information to this Corrigendum.

Supplementary Information is linked to the online version of this Corrigendum at www.nature.com/nature.

naturejobs

THE CAREERS MAGAZINE FOR SCIENTISTS

For many people in the developed world, the Internet has become an extension to their everyday lives. They use it to shop, to get restaurant recommendations and to search out good deals for holidays. In the not-so-distant future, scientists may well be using the web for a spot of shopping and recommendations of their own as they seek the best laboratories to join.

A website is often the first point of contact people have with a laboratory — a sort of digital window into the workings of the place. Most lab websites articulate research programmes and give names and contacts of the lab's members. Good sites also provide ways to seek collaborations. And excellent ones allow potential future members to see what past and present lab participants have gone on to do.

If I was searching for a new lab, I'd want to know that the previous members have published under the principal investigators and have gone on to positions that I'm interested in exploring — whether in academia, industry or government, both on and off the bench. I'd also want to know about lab culture. Do the members interact both professionally and socially? Do they have some sense of humour and a culture of cooperation?

To explore these 'best practices', graduate student, stem-cell scientist and blogger Attila Csordás is hosting a laboratory website competition on his blog (http://pimm.wordpress.com). Csordás' thesis is that few lab websites take full advantage of the medium's technology and don't give visitors the information they want about a group's science and culture. "Am I alone with my opinion that most academic laboratory web pages simply do not meet any advanced, current, dynamic web standards, although this would be crucial for them?" Csordás writes in his blog. He provides a few examples of sites in his discipline that come close, but is challenging life-science labs around the world to share their best efforts. Taking on this challenge will help labs sell their science — and might also attract promising young scientists to their groups.

Paul Smaglik, *Naturejobs* editor

CONTACTS
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock

***Naturejobs* online production:**
Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2–37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

The global challenge

The international effort to address the health crisis in the developing world is providing a wealth of career opportunities. **Virginia Gewin** reports.

G. MENDEL/GLOBAL FUND/CORBIS

The health crisis faced by the developing world is daunting, especially in infectious diseases. According to the World Health Organization (WHO), 8.8 million people were infected with tuberculosis in 2005 and 1.6 million died from the disease, which disproportionately affects the developing world. The HIV pandemic also continues to take its toll in this region. For industrialized nations, helping these countries to tackle such problems is a challenge — but it is one that is attracting increasing funds from governments in the developed world and philanthropic organizations. As a result, international efforts to improve global health now offer a wide range of career opportunities.

The kinds of skills required extend beyond the obvious front-line medical support and can encompass areas such as epidemiology, policy or economics. "Global health is prevention of disease, not simply treatment," says Michael Merson, director of the Global Health Institute at Duke University in Durham, North Carolina.

But for students who wish to pursue a career in global health, the path to success is not always easy to follow. Although a number of institutions have set up training programmes, departments and degrees devoted to the issue, it can still be difficult to find a way into the field. "It is an extremely robust job market for global-health professionals, but the challenge is finding an entry point because there's no defined career track," says Nils Daulaire, president of the Global Health Council, an alliance of global-health organizations based in Washington DC.

Much of the growth in global health has been funded by donations from organizations such as the Bill & Melinda Gates Foundation and the Global Fund to Fight AIDS, Tuberculosis and Malaria. These organizations tend not to be the primary employers, but the initiatives that they fund — to combat disease or generate new vaccines, for example — offer numerous opportunities. The sums of money flowing into the sector are significant. The Gates Foundation has spent some $13 billion since 1994 on global-health programmes, and the Global Fund has provided $7 billion since 2002 to fight the three leading health issues in the developing world.

More than medicine

As a result, there is plenty of research-based work available, from developing AIDS vaccines through to deriving drugs to treat sleeping sickness. But global-health concerns require more than medical intervention. "One of the challenges here is that there are a lot of people entering this area from the clinical and the research side," says Daulaire. "And although both of those are important contributors to the mix, they are not the most critical elements to make a difference to the lives of the 2 billion poorest people in next decade." The skills in demand are management, cultural understanding and an anthropological viewpoint, he adds. Harold Jaffe, a public-health expert at the University of Oxford, UK, agrees, saying that behavioural scientists, economists and anthropologists can "play a role equally important to medical doctors."

Michael Merson: global health is more than treatment.

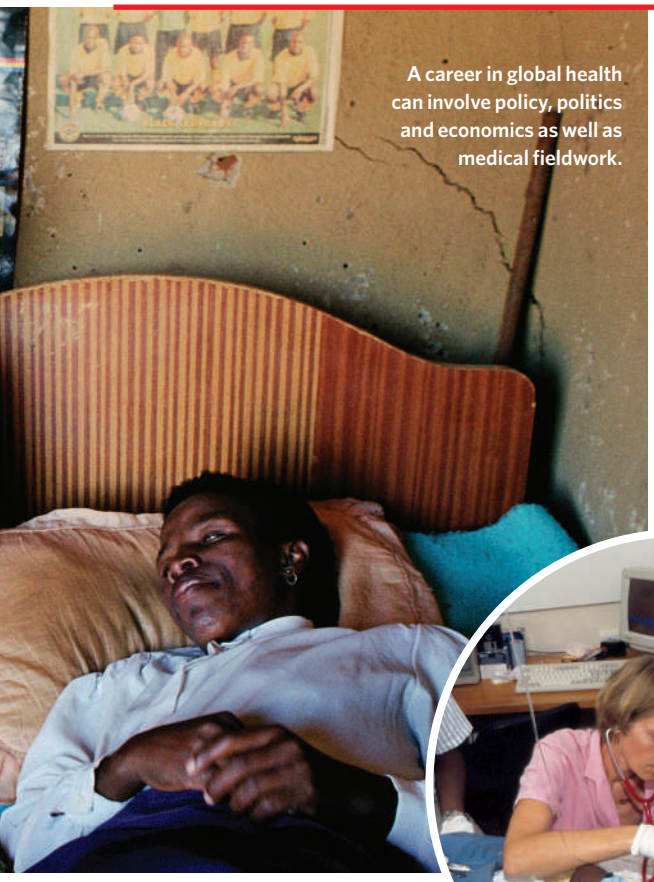
The demand for such a range of skills has contributed to the emergence of health-management organizations, which take skilled scientists and put them to work in a country. Two Boston firms, for example, Management Sciences for Health and John Snow, together employ more than 2,000 workers worldwide focused on every aspect from the clinic to policy development.

And the demand for skills is very real. The US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, for example, is currently recruiting epidemiologists to develop research protocols, to design studies and to track diseases rates. It is also actively looking for public-health advisers to manage health programmes such as immunization activities and AIDS prevention programmes. And it needs laboratory technicians, microbiologists and behavioural scientists to document the social change needed to prevent disease, as well as medical officers to oversee programmes in specific disease areas.

The WHO is looking for similar talents. And, although the focus of the international partnership of the GAVI Alliance is on delivery of vaccines, its needs are also wide-ranging. In addition to epidemiologists, it is increasingly looking for people with skills in programme management and the economics of public-health intervention. It has become increasingly important for non-government organizations and government agencies to monitor the effectiveness of global health programmes to show accountability.

Despite the interest in global health and the intense need for skilled professionals, there is not yet a cohesive career path in developed or developing nations that includes training in the requisite management skills, structured opportunities for overseas experience and well-defined entry-level positions.

"We've realized that training and career development



A career in global health can involve policy, politics and economics as well as medical fieldwork.

BE PREPARED

Top tips for pursuing a career in global health.

- Get overseas experience. Make sure you are cut out for living overseas in difficult conditions with few resources. Programme directors lament the extra costs associated with people abandoning posts because they can't handle the conditions.
- Look for programmes that have well-established partnerships with developing nations.
- Network. Join the Global Health Council or other non-profit groups.
- Seek relevant short courses for your speciality, such as the advanced vaccinology course conducted by Mérieux Foundation and the University of Geneva, designed to educate scientists and biotech decision-makers about preclinical vaccine research as well as design and monitoring strategies for clinical trials.
- Look for mentors who understand the political situation in developing countries.

V.G.

are very fragmented and don't fit together as a proper career path," says Catherine Davies, a populations and public-health-science programme manager at the Wellcome Trust, Britain's largest research charity. To address this, the trust aims to increase its funding to create a series of training fellowships in public health and tropical medicine that should provide a career path starting at the MSc level and leading up to senior posts.

Experience counts

One of the major handicaps for US or European students trying to kick-start a career in global health, is their lack of in-the-field experience. "There is nothing early in one's career as valuable as prolonged field experience to prove you can handle the hardships of living in often difficult situations," says Daulaire. Both the CDC and the WHO offer internships for those hoping to gain overseas experience. Another way to gain experience is to join the US Peace Corps or work with non-profit organizations that are willing to accept volunteers (see 'Be prepared').

Across the United States, the number of training opportunities and faculty positions has increased in recent years. The University of Washington in Seattle, for example, is launching a global-health department that will house 50 faculty members and 500 students over the next few years.

Given the diverse skills in demand, it's not surprising that the training models range from medical to public-health models. The University of California, San Francisco, has created an MS and PhD programme in global health, for instance. And, in collaboration with the University of Washington, the University of Pennsylvania and Johns Hopkins University, it has also set up a programme for clinical scholars that allows residents in eight different disciplines to gain overseas experience through partnerships with Kenya, Uganda, Botswana and Tanzania.

Enhancing opportunities for those who live in developing countries is a priority for many US programmes. In addition to recruiting faculty members, the $110-million Global Health Institute at Emory University in Atlanta, Georgia, has developed a drug-discovery training programme for South Africa. The US National Institutes of Health's Fogarty International Center has 16 programmes for scientists from both the United States and the developing world, most of which are focused on infectious diseases, including AIDS.

Targeted training

In Europe, initiatives such as those sponsored by the Wellcome Trust increasingly try to take training programmes to the countries in question rather than recruiting students from those countries and bringing them to Europe. Norway, for example, has supported the development of master's programmes in international health in conjunction with the Muhimbili University College of Health Sciences in Dar es Salaam, Tanzania.

Politics invariably plays a major role in global health, and it has become important for public-health students to learn the skills necessary to operate in highly politicized environments. Kelley Lee, a reader in global health at the London School of Hygiene and Tropical Medicine, says that students need to be more politically savvy to work with or for the dominant multilateral organizations, such as the WHO, World Bank and World Trade Organization. In an effort to capitalize on multidisciplinary expertise, a collaboration between the London School of Hygiene and Tropical Medicine and the London School of Economics allows MSc students to take courses at both institutions.

Such initiatives may help to address the skills shortage in the sector. But these formal programmes don't come close to bridging the gap between scientific and clinical knowledge in the developed world and local need in the developing one.

Success may, ironically, put those trained in global health out of work, as the point of programmes funded by the Gates Foundation and the Global Fund is to put local people in the developing world in charge of their own public health. But the problem is too large and complicated to make global-health professionals obsolete any time soon. ■

Nils Daulaire worries that career paths are ill-defined.

Virginia Gewin is a freelance science writer based in Portland, Oregon.

NATURE|Vol 447|17 May 2007

MOVERS

Tsuyoshi Kimura, professor, Graduate School of Engineering Science, Osaka University, Japan

2005–07: Technical staff member, Bell Laboratories, Lucent Technologies, New Jersey
2003–05: Technical staff member, Los Alamos National Laboratories, New Mexico

As an undergraduate chemist in 1993, Tsuyoshi Kimura spent an inspiring six months away from his native Japan as an intern at Bell Laboratories in New Jersey. Working with so many pioneering materials scientists convinced the young student that he wanted to follow in their footsteps.

After completing his PhD on superconductivity at Tokyo University in 1996, he worked as a postdoctoral fellow at the Joint Research Center for Atom Technology in Tsukuba. There his projects focused on magnetic transport in correlated electron systems. Working with people interested in patenting discoveries and seeking novel products taught him the importance of industry in his field of research. "The ultimate goal of our research is always application," he says.

Ten years after his first stint in the United States, Kimura was convinced to return, this time to set up his own independent research project. The lure was a combination of academic freedom and the resources available at top-end US institutions. He spent a couple of years at Los Alamos National Laboratories in New Mexico, and then returned to Bell Laboratories, this time staying for two years. It was a fertile and busy period for him: he produced more than 30 publications and won a few prizes, including the Sir Martin Wood Prize in 2005.

"He is one of the most talented materials scientists in the world right now," says John Sarrao, division leader of the Los Alamos materials-physics section. "We were certainly sad to see him go."

But staying in industry was never Kimura's long-term plan. "I always wanted to go back to Japan and actively take part in guiding and educating students," he says. So he was pleased to accept a professorship at Osaka University where he started work on 1 April.

The timing was right, Kimura says. Japan has become a major player in materials sciences. Kimura, whose expertise lies in the synthesis and analysis of condensed matter with superconducting and magnetic behaviour, expects to find a healthy research environment. The university has secured his funding for at least his first two years.

Kimura says that at heart he remains a scientist who is driven by curiosity. "At a university, the process of research is more important than immediate application," he says. "I am looking forward to trying more complicated, adventurous topics." ■

Magdalena Wutte

NETWORKS & SUPPORT

Practical applications

My postdoctoral experience — studying brain–machine interfaces (BMIs) — was pretty good at first. In two years I had collected a lot of human brain data and published a high-impact paper. But the day-to-day routine of sitting in front of a computer analysing data quickly turned to drudgery. Work felt insular and lonely, and I realized that the thrill of scientific discovery no longer motivated me. What I needed was to feel that my work was helping people.

Although I loved being involved in the scientific enterprise, I knew I needed to work on challenging projects in a team setting, and that I was particularly good at technical communication and organization. Focusing on these attributes I came up with a shortlist of potential career options: science policy, science administration and project management.

It was a hard move to make — I had a gnawing fear that accepting a job away from academia was tantamount to selling my soul. Fortunately, my adviser was very supportive, and even offered me a leave of absence to pursue other opportunities.

With his support I investigated several possibilities: advising members of Congress in Washington DC on science-policy issues, administrative positions at a science-funding agency and a scientific centre of excellence, and project-management positions in industry.

But then my adviser won a large contract from the Defense Advanced Research Projects Agency (DARPA) to develop a cortical interface to allow amputees and paralysed individuals to control prosthetic arms using their thoughts. This contract required the lab to build a BMI device and get formal approval to test the device in human clinical trials. The project had everything I was looking for. So I convinced my adviser that he needed a project manager to navigate the regulatory process and manage development activities across institutions and cross-disciplinary teams. Then I convinced him that I was the best person for the job.

A year-and-a-half later and I definitely don't regret my decision to jump off the academic tenure track. The job has been immensely challenging and rewarding, and much more fun than I'd ever imagined. The DARPA contract runs out soon, but I've gained a wealth of experience and I've built a compelling CV that will help me find my next position. ■

Daniel Rizzuto is project manager for the neural prosthetics group at the California Institute of Technology.

POSTDOC JOURNAL

Collegial science

Why are we interested in some things and not others? Why do our interests wax and wane over time? I couldn't help pondering these questions when my adviser left the lab recently for a two-week trip. During his absence, I was surprised to find that sometimes I had difficulty staying focused on my research. As ours is a small lab of just three people (my adviser included), his absence meant a lot of solitary sitting and thinking. The isolation was accentuated by the fact that I'm the only one working on my project.

Solitary science may work for some, but for many, a collaborative research environment seems to be best. In my case, as a newcomer to a vast and established field of study — where the volume of information to sort through and synthesize can seem overwhelming at times — it helps to have colleagues close by who are interested in the same things and who get excited about new findings.

Often, things we think might be interesting turn out to be rather dull, whereas topics we think we're not too keen on can turn into all-consuming passions. A collegial and interactive setting can go a long way towards piquing our interests when we might otherwise struggle to build excitement on our own. Passion for a subject can be contagious, but it's hard to catch anything in isolation. ■

Peter Jordan is a visiting fellow at the National Institute of Diabetes and Digestive and Kidney Diseases in Bethesda, Maryland.

The inside track from academia and industry

All things being equal

Germany now has a law to eliminate discrimination — but interpreting it is proving to be a challenge.

Jens-Peter Mayer

Last year, Germany finally made a legal commitment to abolishing discrimination in the workplace. After years of debate, the General Act on Equal Treatment came into effect in August. Under the terms of this act, it is now illegal to discriminate against job applicants or employees on the grounds of age, race, religion, disability, gender or sexual orientation.

This is a welcome and long overdue step towards genuine equality for the country's workforce. But my experience as a scientific recruitment consultant suggests that there is some confusion, if not downright discomfort, on how to interpret these laws — both from the employer's and the jobseeker's perspective. The fight against inequality requires more than a legal framework. It needs a complete change in the mindsets of both employers and, more broadly, society. The passing of the law in Germany has already sparked greater public debate of the issues, which can only be a good thing.

At a corporate level, employers and recruitment consultants need to review their approach to job candidates to ensure that potentially discriminatory criteria are not being applied. Certain attributes of a candidate have no bearing on whether or not the applicant can do the job and so should not be taken into account. For example, the question of whether a female candidate is likely to want to start a family in the near future is wholly irrelevant to her ability to do the job. But some hiring managers may subconsciously consider this, even if they provide other reasons.

"To reduce discrimination in Germany's workplaces, we need to adapt the law into clear employment practices."

The crux of the problem is that the interpretation of discriminatory actions can be difficult from both ends. For instance, a company could justify not hiring a female employee on the grounds of qualifications. But that rejected applicant could challenge the decision as discriminatory, if she can prove that she is as qualified as her competitors, or if she can prove that the company has an established record of not hiring women of child-bearing age.

Although the new law is designed to halt such discrimination, it may also have some negative effects. Already some job applicants in Germany are trying to twist the new rules to their advantage by applying for jobs they don't want so that they can file legal action once they are rejected. Only a handful of such suits has so far been filed, and none has yet been through the court system.

But the passage of the law did spark some lawsuits, including from three Lufthansa plane pilots, who challenged mandatory retirement ages, and six female executives at a US division of Dresdner Bank who are suing because they were unable to break through a 'glass ceiling' and receive promotions to higher management positions.

These sorts of lawsuits have had a chilling effect on recruiters, who feel they need to exercise a lot more caution. They need to anticipate whether anything they say or any decision they make will result in a lawsuit, as the courts have yet to interpret the new legislation.

As a result, I have heard that some prospective employers are becoming less willing to provide feedback to unsuccessful job applicants. Worried that they might say something that could be interpreted as discriminatory, they are opting to say nothing, which robs jobseekers of potentially constructive criticism.

Recruitment consultants face similar issues. Although we can give personal feedback more easily because people are not always interviewed for a specific vacancy, at Kelly Scientific Resources, we feel we need to reduce the amount of information communicated and must exercise more caution in the way in which we document our candidate evaluations. The way in which we coach individuals with detailed feedback of their performance in an interview must be modified.

"Some prospective employers are becoming less willing to provide feedback to unsuccessful job applicants — worried that they might say something that could be interpreted as discriminatory."

We need to change our practice to protect both ourselves from lawsuits and potential employees from discrimination.

To reduce discrimination in Germany's workplaces, we need to adapt the law into clear employment practices — something that we and many other employers in Germany are now grappling with. Discrimination is clearly unacceptable. But is there truly no discrimination in countries with anti-discrimination laws? Job statistics related to race or culture suggest that there is still a gap between the desired outcome and reality. Legislating over discrimination is a first step in closing that gap. Interpreting those laws in the workplace is the next. But the biggest leap — making sure these laws are clear to understand and implement for both sides — will be necessary to eliminate it. ■

Jens-Peter Mayer is branch manager for Kelly Scientific Resources in Cologne, Germany.